Florin Alin Sava este cadru didactic titular al Universității de Vest din Timişoara şi doctorand în psihologie la Universitatea „Babeş-Bolyai" din Cluj-Napoca (conducător ştiinţific Prof. univ. dr. Nicolae Jurcău). A efectuat o specializare în statistică avansată la Universitatea Ann Arbor din Michigan, SUA. Domeniile de interes sunt metodologia cercetării şi statistică, psihologia cognitivă şi psihologia personalității. În prezent este titularul disciplinei „Analiza superioară a datelor" la specializarea Psihologie şi al laboratoarelor de Statistică de la secțiile Sociologie şi Asistență Socială. A realizat primul site de statistică pentru ştiințele sociale în limba română, disponibil la adresa http://statisticasociala.tripod.com. A participat în 2001, ca lector invitat, la Şcoala Națională de Statistică Socială organizată la Cluj-Napoca de Fundația pentru o Societate Deschisă, cu o prelegere vizând regresia logistică, iar în 2002, la Şcoala de Vară pe Psihologie Organizațională organizată de Asociația de Psihologie Industrială şi Organizațională din România, cu o prezentare vizând scalele multidimensionale.

Când vedem o carte de statistică psihologică nu ne aşteptăm să găsim prea multe lucruri noi. Aceleaşi formule, aceiaşi algoritmi de calcul, eventual utilizând alte simboluri şi poate alte exemple. Puține sunt lucrările din domeniul amintit care să ne învețe ***de ce folosim o anumită formulă şi nu alta, cum să interpretăm rezultatele obținute*** *şi* ***care este procedura de utilizare a calculatorului*** *pentru a surprinde ceea ce dorim să aflăm sau să verificăm. Cartea pe care Florin Sava o oferă utilizatorilor de statistică psihologică şi educațională rezolvă multe din problemele cu care se confruntă aceştia. Poate că o calitate a acestei lucrări este aceea că este scrisă de către un psiholog. El nu dă nişte "rețete" lipsite de explicații, abundente în formule şi demonstrații redundante. Această carte este fundamentală pentru un psiholog sau student prin aceea că îi spune clar* ***ce şi de ce să facă într-un anumit mod şi nu în altul.***

Prof. univ. dr. Horia D. Pitariu
Universitatea Babeş-Bolyai, Cluj-Napoca

H.G. Wells spunea: "Într-o zi, gândirea statistică va fi la fel de necesară oricărui cetățean folositor societății, ca scrisul şi cititul". Autorul prezentei lucrări continuă această idee, oferind celor interesați (studenți, masteranzi, doctoranzi), alfabetul descifrării statisticii aplicate în psihologie.

Conf. univ. dr. Mihai Hohn
Universitatea de Vest din Timişoara

ISBN 973-7973-11-9

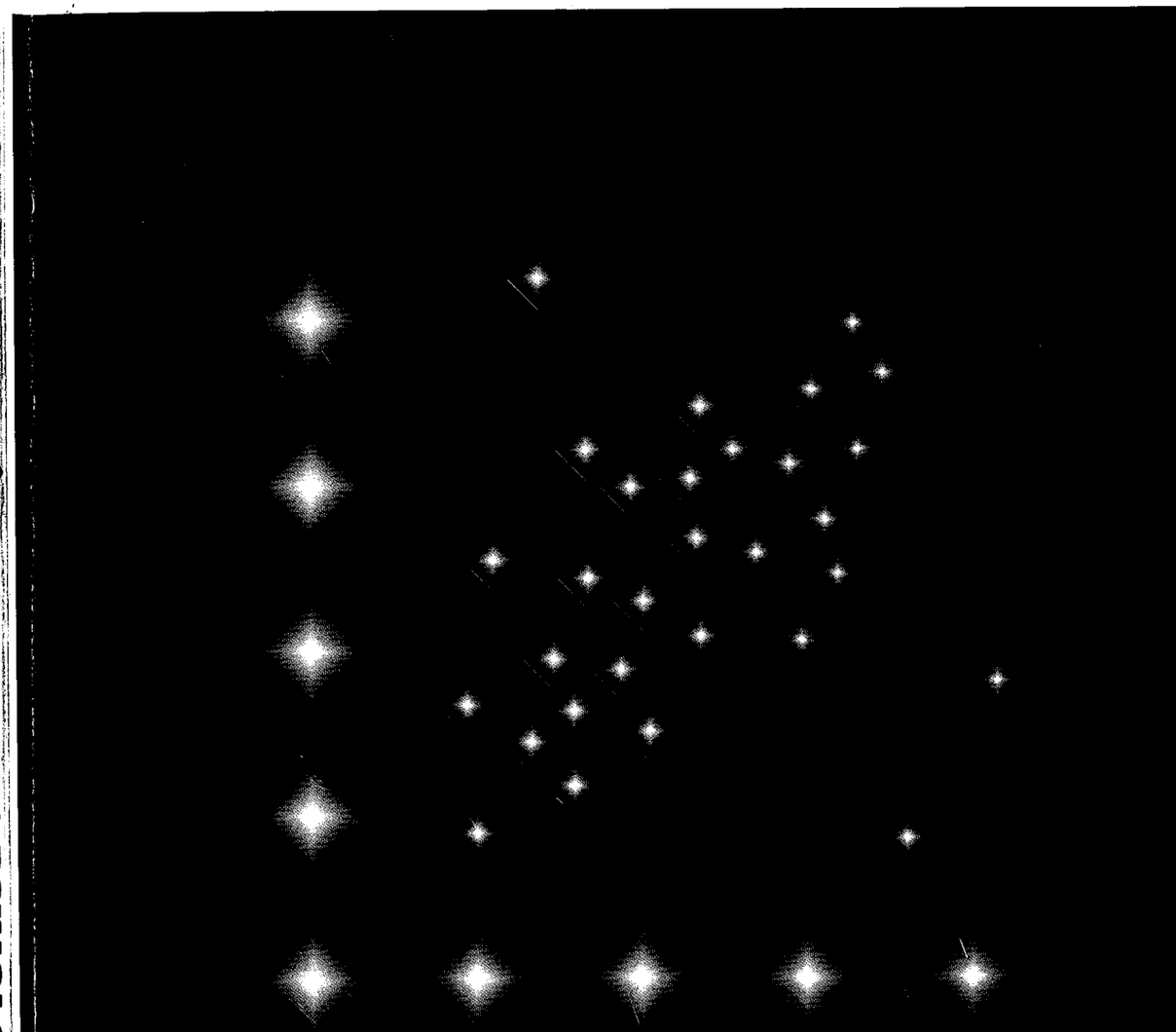
Florin Sava

ANALIZA DATELOR ÎN CERCETAREA PSIHOLOGICĂ
Metode statistice complementare

Florin Sava

ANALIZA DATELOR ÎN CERCETAREA PSIHOLOGICĂ

Metode statistice complementare



Editura ASCR

Florin Sava

**ANALIZA DATELOR ÎN CERCETAREA PSIHOLOGICĂ.
METODE STATISTICE COMPLEMENTARE**

Copyright © 2004 Asociația de Ştiințe Cognitive din România
Toate drepturile rezervate. Reproducerea integrală sau parțială a textului fără acordul prealabil în scris al editurii este interzisă.

Procesare computerizată şi coperta: Daniel Paul

Corectura: Florentina Frățilă

Editura ASCR (Asociația de Ştiințe Cognitive din România)
Str. Gh. Bilaşcu nr. 37, Cluj-Napoca
Tel.: (+40)-264-590967
Fax: (+40)-264-595576
Email: ascr@psychology.ro

Pentru comenzi:
Tel.: (+40)-264-590967 sau 0744.798543
Email: ascr@psychology.ro

ISBN 973-7973-11-9

FLORIN SAVA

ANALIZA DATELOR ÎN CERCETAREA PSIHOLOGICĂ. METODE STATISTICE COMPLEMENTARE

Editura ASCR
Cluj-Napoca
2004

Părinților mei, Viorel şi Zorița,
pentru că m-au învățat să privesc lumea fără ură.

CUPRINS

PREFAȚĂ ix

INTRODUCERE xi

Capitolul 1. ELEMENTE DE STATISTICĂ DESCRIPTIVĂ 1

1.1. Modalități de măsurare a variabilelor studiate 2
1.2. Analiza descriptivă univariată 3
1.3. Analiza descriptivă bivariată 12
1.3.1. Coeficienții de corelație 13
1.3.2. Reprezentarea grafică a asocierii dintre variabile 15
1.3.3. Tabele de contingență – moduri de prezentare a asocierii dintre datele non-numerice 24
Bibliografie 26

Capitolul 2. TESTAREA IPOTEZELOR 27

2.1. Ipoteza nulă – clarificări conceptuale 28
2.1.1. Logica testării ipotezelor 29
2.1.2. Ipoteza nulă şi mărimea eşantionului 31
2.1.3. Erorile de tip I şi II în testarea ipotezelor 31
2.2. Mărimea efectului 33
2.2.1. Indicatori ai mărimii efectului 34
2.2.2. Interpretarea mărimii efectului 35
2.2.3. Stabilirea mărimii efectului 37
2.2.4. Limitele mărimii efectului 37
2.3. Puterea statistică a unei cercetări 38
2.4. Appendix – Ce înseamnă gradele de libertate? 46
Bibliografie 48

Capitolul 3. TEHNICI INTRODUCTIVE DE TESTARE A IPOTEZELOR 51

3.1. Testarea asocierii dintre două variabile 51
3.1.1. Coeficientul de corelație r 51
3.1.2. Corelația parțială 59
3.2. Tehnici de comparație între grupuri 61
3.2.1. Tehnicile χ^2 61
3.2.2. Testele t şi z de comparare a diferențelor dintre două medii 70
3.3. Managementul datelor brute 81
3.3.1. Managementul datelor lipsă 81
3.3.2. Managementul datelor neobişnuite 82
3.3.3. Transformarea datelor brute 82
Bibliografie 85

Capitolul 4. TEHNICILE ANOVA **87**

4.1. ANOVA simplă 90
4.1.1. Procedurile post hoc 92
4.1.1.1. Mărimea efectului în cadrul procedurilor post hoc 95
4.1.2. Procedurile a priori 97
4.1.2.1. Contraste nestandardizate 97
4.1.2.2. Mărimea efectului în cazul contrastelor 101
4.1.2.3. Contrastele standardizate 102
4.1.2.4. Contrastele polinomiale 104
4.2. ANOVA factorială 107
4.2.1. Metoda efectelor simple 111
4.2.2. Mărimea efectelor în cazul ANOVA factorială 114
4.2.3. Condiții necesare pentru aplicarea ANOVA simplă şi ANOVA factorială 115
4.3. ANOVA cu măsurători repetate 116
4.3.1. Mărimea efectelor în ANOVA cu măsurători repetate 121
4.3.2. Condiții necesare pentru aplicarea ANOVA cu măsurători repetate 121
4.4. Alte modalități ANOVA 124
4.5. Exemplu de ANOVA factorială: cazul euristicii de accesibilitate 125
4.6. Tehnicile ANCOVA 128
4.6.1. Mărimea efectului în tehnicile ANCOVA 132
4.6.2. Condiții suplimentare aplicării ANCOVA 133
4.7. Tipuri de factori în design-urile de tip ANOVA sau ANCOVA 135
Bibliografie 137

Capitolul 5. ANALIZA FACTORIALĂ **139**

5.1. Analiza factorială exploratorie 140
5.1.1. Teoria analizei factoriale 141
5.1.2. Analiza componentelor principale vs. analiza factorială 143
5.1.3. Design-ul şi metodologia analizei factoriale 145
5.1.3.1. Tipuri de matrice şi metode de extragere a factorilor 146
5.1.3.2. Metode de selectare a numărului adecvat de factori 148
5.1.3.3. Analiza statistică a factorilor extraşi 153
5.1.3.4. Rotirea factorilor 155
5.1.3.5. Interpretarea naturii factorilor 160
5.1.3.6. Ameliorarea structurii factoriale 161
5.1.3.7. Aspecte metodologice complementare 163
5.1.4. Exemplu de analiză factorială exploratorie - cazul inteligenței 167
5.2. Analiza factorială confirmatorie 170
5.2.1. Specificarea modelelor concurente 171
5.2.2. Identificarea modelelor 173
5.2.3. Stabilirea gradului de adecvare a modelelor 176
5.2.3.1. Indicatori absoluți 176
5.2.3.2. Indicatori de comparare 178
5.2.4. Modificarea modelului 180
5.2.5. Aspecte metodologice ale aplicării analizei factoriale confirmatorii 182
Bibliografie 183

Capitolul 6. REGRESIA LINIARĂ **185**

6.1. Regresia liniară simplă 187
6.1.1. Principii ale regresiei liniare 187
6.1.2. Evaluarea eficienței unei ecuații de regresie 192
6.1.3. Testarea individuală a eficienței predictorilor în estimarea evoluției criteriului 194
6.2. Modalități de analiză în regresia multiliniară 197
6.3. Regresia multilinară în scop predictiv 200
6.3.1. Interpretarea coeficienților de regresie 204
6.3.2. Rolul ecuației de regresie în interpretarea datelor cu scop predictiv 208
6.3.3. Instrumente complementare regresiei în scop predictiv 209
6.3.3.1. Decizia organizațională 209
6.3.3.2. Contravalidarea 212
6.3.3.3. Analiza cazurilor influente 214
6.4. Regresia multiliniară în scop explicativ 216
6.4.1. Logica regresiei utilizate în scop explicativ 217
6.4.2. Testarea eficienței explicative a modelului propus 219
6.4.3. Interpretarea coeficienților de regresie în analizele cu scop explicativ 220
6.4.4. Indicatorii mărimii efectului în cazul regresiei multiple 222
6.5. Condiții necesare pentru aplicarea regresiei liniare 223
6.5.1. Evitarea erorii de specificare 223
6.5.2. Evitarea erorii de măsurare 224
6.5.3. Eroarea reziduală cât mai mică 225
6.5.4. Evitarea multicoliniarității 225
6.5.5. Existența unor relații de tip liniar şi aditiv 227
6.6. Exemplu de regresie multiplă liniară – cazul citirii 228
6.7. Direcții de evoluție ale regresiei 232
6.7.1. Alte tipuri de regresie 232
6.7.2. Identificarea efectelor de interacțiune în regresia multiplă 232
6.7.3. Analiza de cale 238
Bibliografie 241

Capitolul 7. REGRESIA LOGISTICĂ **243**

7.1. Concepte şi principii teoretice 243
7.2. Modelul regresiei logistice 245
7.3. Estimarea modelelor logistice 249
7.3.1. Testarea modelelor explicative logistice 250
7.3.2. Compararea modelelor logistice 253
7.3.3. Stabilirea eficienței explicative a modelului ales 256
7.3.4. Testarea şi interpretarea coeficienților logit 260
7.3.5. Stabilirea unor profile ideale 264
7.3.6. Analiza cazurilor influente asupra modelului 265
7.4. Exemplu de regresie logistică binară – cazul infidelității 266
Bibliografie 269

Capitolul 8. SCALAREA MULTIDIMENSIONALĂ 271

8.1. Modele de scalare multidimensională 272
8.2. Design-ul cercetării în scalarea multidimensională 275
8.2.1. Selectarea obiectelor şi a numărului acestora 276
8.2.2. Metode de colectare a datelor 277
8.2.2.1. Compararea fiecărei perechi de stimuli 277
8.2.2.2. Sortarea stimulilor 278
8.2.2.3. *Ierarhizarea stimulilor* 279
8.2.3. Managementul datelor colectate prin SPSS 279
8.3. Harta perceptivă. Reprezentarea spațială a distanțelor psihologice 283
8.3.1. Selectarea numărului de dimensiuni din harta perceptivă 283
8.3.2. Interpretarea configurației MDS 288
8.3.3. Aplicații INDSCAL 293
8.4. Modele de scalare multidimensională conexe 295
8.4.1. Modele MDS constrânse 295
8.4.2. Modele MDS extinse 296
8.5. Incursiune în opțiunile SPSS 298
Bibliografie 299

PREFAȚĂ

O CARTE DE STATISTICĂ PENTRU PSIHOLOGI

Când mi s-a propus să redactez un cuvânt introductiv la cartea colegului meu, Florin Sava, am fost oarecum şocat, deoarece sunt mai mult un consumator de statistică decât un creator al acestui domeniu. Există şi percepția, uneori falsă, că psihologii nu sunt prea buni statisticieni. Mulți chiar se complac în această convingere şi nu fac nici un efort să-şi optimizeze performanțele într-un domeniu, care, de ce să nu recunoaştem, dă substanță oricărei cercetări din domeniul psihologiei. Personal, nici nu pot concepe o investigație psihologică fără o abordare statistică adecvată, indiferent că se pune problema unui demers cantitativ sau calitativ.

Psihologia este atestată ca o ştiință experimentală. Experimentul a fost acel element care a conferit psihologiei calitatea de ştiință. Ea este evaluată ca fiind o ştiință tânără, născută ca o sinteză a cercetărilor lui E.H. Weber asupra senzațiilor tactile (1849), ale lui H. Helmholtz asupra senzațiilor vizuale şi optice (1862-1867) şi studiilor de psihofizică (1860) ale lui G. Th. Fechner, în laboratorul lui Wilhelm Wundt de la Leipzig, mai ales după publicarea manualului său de psihologie (1874). Aplicarea în mod consecvent a metodei experimentale ca regulă în investigarea psihologică a însemnat renunțarea la sistemele vaste filosofico-psihologice. Psihologia cunoaşte în prezent o inserție totală în activitățile practice în care deseori se reflectă diverse categorii de probleme cu care se confruntă ființa umană în contextul existenței sale cotidiene. În toate aceste cadre acționale, psihologii dezvoltă proiecte experimentale pe care le fundamentează prin intermediul unui aparat statistic construit cu multă grijă. Prin utilizarea experimentului în psihologie s-a pus accent pe colectarea de date obiective privind comportamentul uman, astfel fiind posibilă generarea a numeroase legități privind senzațiile, percepțiile, reprezentările, memoria etc., legități care au permis psihologiei să devină o ştiință independentă. Psihologia ştiințifică este o provocare vizavi de practicile diletante sau abordările parapsihologice de genul astrologiei, psihobioritmurilor, telepatiei, grafologiei, chiromanției etc. Acestor practici cu profund caracter neştiințific, cât şi adepților lor, psihologii le-au răspuns printr-o abordare ştiințifică, proiectând o serie de experimente în care erau folosite proceduri statistice adecvate. La timpul său, psihologul american R. Pintner sublinia față de pretențiile caracterologilor şi ale altor şarlatani, că pe aceştia nu-i poți combate decât cu fapte, prin aceasta înțelegându-se argumente ştiințifice, rezultate ale unor investigații. Ca atare, de-a lungul timpului au fost proiectate şi publicate o serie de investigații menite să demonstreze caracterul neştiințific al astrologiei, horoscopului, psihobioritmurilor etc.

Cercetarea psihologică contemporană a adus statistica psihologică şi educațională la un punct suficient de înalt, încât să nu mai fie descifrată cu prea multă uşurință decât cu un efort susținut şi după o practică îndelungată de către orice cercetător. Azi, problema calculelor, cu care se confruntau cercetătorii până nu de mult, nu mai constituie o barieră, calculatorul şi pachetele de programe statistice făcând posibilă printr-o simplă apăsare pe o tastă, efectuarea celor mai complicate calcule într-un timp nebănuit de scurt. Creatorul cunoscutului chestionar de personalitate 16PF, psihologul R.B. Cattell (1905-1998), s-a folosit de o simplă maşină de calcul pentru a descoperi cei 16 factori ai celebrului său chestionar. Se spune că foile de hârtie pe care erau trecute iterațiile analizei factoriale se întindeau din casă până în grădină, fiind asistat de câteva secretare. Unele voci afirmă chiar că în noianul de calcule s-au strecurat şi unele erori. Cu tehnica actuală de calcul, acest lucru nu se mai poate întâmpla. Un singur calculator, software-ul adecvat şi un operator, pot rezolva orice problemă statistică. Deci, problema nu mai este a dificultății de a realiza şi înțelege demersul de calcul al cutărui sau cutărui indice statistic, ci de a-i

descoperi semnificația, de a-l interpreta corect. Psihologul de azi este prea puțin interesat de unde provin unele formule de calcul, cum sunt ele aplicate din punct de vedere tehnic. El se concentrează pe tipul de statistici utilizate în diferite contexte de cercetare şi pe cum trebuie acestea să fie interpretate. A alege corect abordarea statistică adecvată demersului experimental abordat este într-adevăr o problemă peste care nu se poate trece cu prea multă uşurință. Desigur, nici abuzul de statistică nu este de dorit. Uneori s-a ajuns ca faptul psihologic să fie mascat de o abordare statistică prea laborioasă. S-a ajuns până acolo, încât unele articole din reviste prestigioase de psihologie să fie de neînțeles pentru psihologul obişnuit, ele devenind accesibile numai unui număr restrâns de psihologi. Or, în acest caz, pierde şi revista şi psihologia ca ştiință. Să nu uităm că inspectarea unei simple diagrame de corelație ne spune adesea mai mult decât coeficientul de corelație. Ea ne permite să identificăm unele concentrări specifice ale anumitor scoruri, să intuim elementele distorsionante care pot afecta corelația etc. Apoi, diagrama de corelație ne indică acele cazuri de excepție care pot constitui subiectul unui studiu de caz, cele care se abat de la o regulă generală. Despre statistica psihologică şi educațională se pot spune multe, la fel de multe şi despre cei care sunt incapabili să o practice, afirmând *iresponsabil* că ei "preferă" metodele calitative, neavându-se prea bine cu statistica. Abordarea cantitativă şi calitativă *nu trebuie să fie disociate*, ele se completează reciproc.

Când vedem o carte de statistică psihologică nu ne aşteptăm să găsim prea multe lucruri noi. Aceleaşi formule, aceiaşi algoritmi de calcul, eventual utilizând alte simboluri şi poate alte exemple. Puține sunt lucrările din domeniul amintit care să ne învețe *de ce folosim* o anumită formulă şi nu alta, *cum să interpretăm* rezultatele obținute şi *care este procedura de utilizare a calculatorului* pentru a surprinde ceea ce dorim să aflăm sau să verificăm. Cartea pe care Florin Sava o oferă utilizatorilor de statistică psihologică şi educațională rezolvă multe din problemele cu care se confruntă aceştia. Poate că o calitate a acestei lucrări este aceea că este scrisă de către un psiholog. El nu dă nişte "rețete" lipsite de explicații, abundente în formule şi demonstrații redundante. Această carte este fundamentală pentru un psiholog sau student prin aceea că îi spune clar *ce şi de ce să facă într-un anumit mod şi nu în altul*. Ea oferă informații pertinente despre ceea ce se găseşte în spatele formulelor statistice pe care le utilizăm într-o cercetare. Mai mult, prin numeroasele exemple pe care le găsim la tot pasul în această lucrare, se încearcă o altfel de interpretare a fenomenului psihologic, mai apropiată de realitate şi cu nenumărate oportunități interpretative,

O carte de statistică se studiază cu creionul în mână. Lucrarea lui Florin Alin Sava am citit-o făcând unele note succinte pe manşeta paginii şi în fața calculatorului pentru a urmări, uneori din pură curiozitate, algoritmii de calcul propuşi. În acest context, ea a constituit realmente o provocare. Aş mai menționa o calitate care mi s-a părut importantă: umorul. Nu se poate să studiezi lucrarea în cauză fără să zâmbeşti. Despre statistică se spun multe lucruri comice, dar Florin a ştiut să le distribuie proporțional şi inteligent în cartea sa. Exemplele rezultate din practică sau simplu inventate, ne dau ancore simpatice în penetrarea sensurilor variatelor abordări statistice, indiferent de cât de complicate ar fi ele.

Cartea colegului Florin Sava este dedicată în primul rând studenților în psihologie care, ştim prea bine că rareori fac o pasiune din studiul statisticii, dar care sunt obligați să o utilizeze. Ea este un foarte bun ghid al oricărui cercetător din domeniul psihologiei, al oricărui psiholog pentru care psihologia este o ştiință cu puternice rădăcini aplicative.

Prof. Horia D. Pitariu
Universitatea "Babeş-Bolyai"
Cluj-Napoca

INTRODUCERE

Psihologia românească a făcut paşi importanți pentru a recupera pierderile provocate de o decizie politică neinspirată a regimului comunist. Se vorbeşte tot mai mult de o psihologie autohtonă care să nu mai fie tributară unor practici imitative, de divinizare a psihologiei americane. Cu siguranță, această perspectivă meritã încurajată pentru a oferi psihologiei românești un stil propriu.

Totuşi, există destul de mulți psihologi care confundă acest deziderat cu o larghețe metodologică greşit înțeleasă. Erorile metodologice sau ignoranța statistică sunt, din nefericire, prezente frecvent în practica psihologilor de la noi. Dacă subiectele de cercetare specifice spațiului social românesc merită promovate, standardele ştiințifice de ordin metodologic şi statistic trebuie păstrate la un nivel ridicat, cel puțin similar cu cel întâlnit în majoritatea publicațiilor de specialitate de limbă engleză.

Pentru a realiza acest lucru este nevoie acută de texte statistice comprehensibile. În ultimii ani, această cerere a început să fie acoperită prin contribuțiile editoriale, în special ale colegilor din Iaşi sau Cluj-Napoca. Lucrarea de față se alătură acestor demersuri, oferind o viziune proprie asupra statisticii, ce are ca o primă sursă stagiul de specializare de la Universitatea Michigan din Ann Arbor din 2000. Însă sursa principală de inspirație în realizarea acestei cărți are în vedere întrebările şi sugestiile studenților de la Facultatea de Sociologie şi Psihologie, specializarea Psihologie, din cadrul Universității de Vest din Timişoara şi discuțiile profesionale cu colegii. Lor le datorez în cea mai mare parte acest demers intelectual.

Bunăoară, formarea predominant umanistă a majorității studenților de la psihologie îi îndepărtează, cel puțin în plan afectiv, de o abordare matematică a statisticii. De aceea, pe întreg parcursul lucrării am încercat să evit, pe cât posibil, un asemenea demers, substituindu-l printr-o prezentare a logicii acestor tehnici şi printr-o multitudine de exemple-suport.

Cartea se adresează studenților interesați, doctoranzilor şi tinerilor cercetători care urmăresc să-şi clarifice multe din conceptele statistice vehiculate. Din punctul de vedere al gradului de complexitate, lucrarea are un nivel intermediar de dificultate, fiind deosebit de utilă mai ales după parcurgerea unui curs introductiv de statistică.

Din punctul de vedere al conținuturilor vehiculate, cartea este structurată pe trei părți, criteriul de divizare fiind dat de complexitatea cunoştințelor pe care trebuie să le dețină cititorul. Astfel, partea I este alcătuită din trei capitole introductive vizând statistica descriptivă, testarea ipotezelor şi tehnicile t, χ^2 sau r, ca modalități uzuale de testare a ipotezelor. Această secțiune constituie, în principal, o recapitulare originală a cunoştințelor acumulate de cititori în urma unui

curs introductiv de statistică. Secțiunea nu poate fi neglijată de către cei care dispun de cunoştințe lacunare în domeniu, parcurgerea celor trei capitole fiind necesară pentru a putea înțelege celelalte părți ale cărții.

Partea a II-a, cea mai voluminoasă de altfel, cuprinde tehnicile statistice cele mai utilizate în cercetarea psihologică de vârf: regresia multiliniară, tehnicile ANOVA sau ANCOVA, analiza factorială exploratorie sau cea confirmatorie. În privința lor există o situație paradoxală. Deşi majoritatea cursurilor introductive de statistică nu ajung să cuprindă aceste tehnici din cauza lipsei de timp, cele mai multe articolele de specialitate publicate utilizează aceste tehnici, mai ales în literatura de limbă engleză. Prin urmare, multe persoane întâmpină dificultăți în înțelegerea demersului statistic din cadrul acestor cercetări, fiind nevoite să studieze individual ori să preia fără o reflecție critică orice studiu publicat. Din nefericire, în literatura de limbă română există puține texte care prezintă adecvat aceste subiecte, cele mai multe tratând problemele într-o manieră nepotrivită pentru profilul studentului sau absolventului de psihologie. Lucrarea de față îşi propune să depăşească aceste neajunsuri, obiectivul central al părții a II-a fiind „traducerea" metodelor statistice utilizate frecvent în cercetare pe înțelesul psihologilor.

Pentru a realiza acest lucru, demersul matematic de tip hârtie-creion a fost substituit prin prezentarea şi discutarea rezultatelor oferite de două din cele mai întâlnite programe statistice: SPSS, în varianta 11.0 şi AMOS, în varianta 4.0. Pe lângă interpretarea rezultatelor oferite de aceste programe, fiecare capitol conține una sau mai multe baze de date care pot fi descărcate gratuit de pe internet la adresa http://statisticasociala.tripod.com. Fişierele respective pot fi utilizate de cititorii care dispun de aceste programe statistice pentru a exersa problemele discutate.

În ultima parte a cărții sunt prezentate două metode statistice complementare: scalarea multidimensională şi regresia logistică. Ambele metode sunt considerate avansate din punctul de vedere al gradului de complexitate, pentru a le putea înțelege fiind nevoie de o bună stăpânire a primelor două secțiuni ale cărții. De pildă, înțelegerea capitolului destinat regresiei logistice necesită cunoaşterea regresiei liniare, în timp ce scalarea multidimensională face apel atât la informațiile vizând regresia liniară, cât şi la cele găsite în paginile referitoare la analiza factorială.

În încheiere, aş dori să mulțumesc tuturor celor care au contribuit prin sugestiile şi observațiile lor la bunul mers al acestei lucrări. Aceleaşi aprecieri deosebite vi le urez şi dumneavoastră, dragi cititori, pentru osteneala şi curajul de a da piept cu un limbaj statistic pe alocuri neprietenos, cât şi pentru eventualele sugestii la afsava@socio.uvt.ro. Lectură utilă!

Timişoara, 25.01.2004

Autorul

CAPITOLUL 1

ELEMENTE DE STATISTICĂ DESCRIPTIVĂ

Analiza datelor poate părea unora un proces anost, rutinier. Însă importanța acestui proces a fost sesizată de către întreaga comunitate ştiințifică. Până şi titulatura de cercetare ştiințifică cantitativă sau calitativă are în vedere tocmai modul de analiză a datelor.

În prezent, comunitatea psihologilor recunoaşte şi încurajează ambele tipuri de cercetare, cu toate că, în unele medii conservatoare, cercetarea cantitativă are întâietate. Acest fapt poate fi regăsit şi în practica unor cadre didactice universitare care încurajează întotdeauna aplicarea unor metode cantitative complementare unui demers cantitativ, în schimb, utilizează cu caracter facultativ recomandarea unor demersuri calitative într-o cercetare cantitativă.

Cartea de față vizează doar tehnici statistice cantitative, însă nu face acest lucru dintr-o lipsă de respect față de metodele calitative. Tehnicile cantitative alese pentru a fi descrise vin să completeze metodele prezentate în textele introductive din domeniu, precum cele ale lui Clocotici şi Stan (2000), Novak (1977), Hohn (2000), Radu şi colab. (1993), Rotariu şi colab. (1999). Aşa cum am precizat şi în introducere, deşi tehnicile prezentate în lucrarea de față sunt foarte răspândite în articolele ştiințifice publicate în revistele de specialitate, cursurile introductive de statistică nu ating decât tangențial această tematică. Prin urmare, studenții şi cercetătorii din psihologie se văd nevoiți, de multe ori, să accepte într-un mod necritic rezultatele publicate sau prezentate la diverse comunicări, din cauza lipsei de cunoştințe în domeniu.

Analiza datelor cu scop descriptiv include un complex de metode şi tehnici menite să faciliteze înțelegerea realității prin analiza cantităților observate. Reprezentările grafice, mediile, tabelele de frecvențe sau alți indicatori statistici reprezintă elemente întâlnite des în cotidian sau în textele de specialitate, de aceea este necesară o bună cunoaştere a lor.

Acest minim de cunoştințe poate fi integrat în termenul generic de **statistică descriptivă.** Ea nu-şi propune altceva decât să facă o radiografie cât mai corectă a realității. Dacă analizăm un singur aspect al realității vorbim despre o **statistică descriptivă univariată** (pentru o singură variabilă), iar în situația în care descriem evoluția a două variabile, vorbim de o **statistică descriptivă bivariată** (pentru două variabile). Se vorbeşte şi despre o **statistică multivariată** în situațiile în care sunt analizate simultan cel puțin trei aspecte ale realității. Prelucrarea statistică ține cont nu doar de numărul variabilelor, ci şi de calitatea scalelor de măsurare utilizate.

O altă idee de bază în analiza statistică a datelor are în vedere diferența dintre **analiza la nivelul lotului** testat şi **analiza la nivelul populației** din care face parte acel eşantion. De exemplu, putem obține anumite valori descriptive la nivelul lotului testat (ex. nivelul de empatie la o clasă de elevi de clasa a IX-a) şi alte valori la nivelul populației (ex. valoarea medie a empatiei la nivelul populației de elevi de clasa a IX-a). De multe ori, analiza datelor se bazează pe rezultatele obținute la nivelul lotului de subiecți testați, însă inferența se face la nivelul populației din care acesta face parte. Pentru a facilita distincția dintre cele două domenii se utilizează termeni diferiți precum cei de parametri, respectiv indicatori statistici. Parametrii se referă la nivelul populației, iar indicatorii statistici se referă la nivelul lotului analizat. Această diferențiere se regăseşte, frecvent, şi în modul de notare. Astfel, media şi abaterea standard la nivelul populației se notează cu μ, respectiv σ, iar la nivelul eşantionului se utilizează simbolurile m, respectiv s.

Vom reaminti, pe scurt, informațiile de bază din statistica descriptivă, pentru a pătrunde mai bine semnificația informațiilor prezentate în capitolele următoare. Cititorii interesați pot consulta lucrările deja menționate pentru mai multe detalii.

Acestea vor fi grupate în trei secțiuni: (1) importanța modului de măsurare a variabilelor în analiza datelor; (2) analiza descriptivă univariată şi (3)analiza descriptivă bivariată.

1.1. Modalități de măsurare a variabilelor studiate

Primul element de bază în analiza datelor este identificarea modului în care au fost măsurate variabilele de interes. Astfel, se pot distinge patru scale de măsurare: nominală, ordinală, de interval şi proporții (raport), fiecare având calități diferite în ceea ce priveşte posibilitățile pe care le oferă pentru analiza datelor.

De pildă, variabilele de tip **nominal** presupun răspunsuri care nu sunt numerice şi nici nu pot fi ierarhizate. Prin urmare, scalele de tip nominal implică analize de frecvențe. Un exemplu tipic de variabilă măsurată prin scală nominală este „apartenența la religie" deoarece răspunsurile aferente – ortodox, catolic, protestant, musulman etc. – nu sunt nici numere, nici ordonabile intrinsec. Variabilele de tip **ordinal**, deşi nu sunt de tip numeric, permit ordonarea intrinsecă a răspunsurilor, permițând astfel realizarea unor ierarhii. Acest fapt permite cercetătorilor să apeleze la analize bazate pe ierarhii (ranguri). Un exemplu de variabilă măsurată prin scală ordinală este întâlnit frecvent în studiile despre atitudini şi implică răspunsuri de genul „foarte rar", „rar", „mediu", „des", „foarte des". Un alt tip de scală de măsurare este cea de tip **interval**. Aceasta este o scală numerică (ex. un IQ de 120; 7 puncte la o probă de asertivitate etc.). Tot o scală numerică este şi scala de **proporții**, avantajul acesteia față de scala de tip interval constând în faptul că valoarea zero nu este aleasă arbitrar. Un exemplu de variabilă măsurată prin scale de proporții este timpul de reacție (ex. 170 ms) sau

venitul lunar (ex. 10.000.000 lei). Unii autori, precum Kinnear şi Gray (2000), tind să restrângă ultimele două modalități de măsurare la termenul generic de scală numerică deoarece majoritatea tehnicilor statistice utilizate în cazul scalelor de tip interval sunt valabile şi în cazul scalelor de tip proporții. Aceeaşi poziție a fost adoptată şi de consilierii statistici ai SPSS-ului, care au renunțat la a mai diferenția între cele două tipuri de variabile (SPSS 11.0, 2001).

Un alt aspect ce trebuie subliniat cu privire la scalele de măsurare este acela că aproape orice variabilă poate fi măsurată prin mai multe scale de măsurare. Spre exemplu, nivelul de şcolarizare poate fi evaluat prin scală nominală în cazul unor răspunsuri de tip „şcolarizat" sau „neşcolarizat" sau prin scală ordinală, în cazul unor răspunsuri precum „studii primare", „studii medii", „studii postliceale", „studii universitare" etc. Realizarea unor scoruri vizând gradul de alfabetizare sau nivelul cunoştințelor generale implică obținerea unor indicatori numerici, specifici scalelor de tip interval. În cele din urmă, prin intermediul scalei de proporții, nivelul de şcolarizare poate fi reflectat prin numărul de ani de şcoală absolviți (ex. 5; 12; 16 etc.).

De asemenea, trebuie remarcată distincția calitativă dintre scalele nominale şi ordinale, pe de o parte, respectiv scalele de interval şi proporții, pe de altă parte. Primele sunt scale non-numerice, presupunând apelul la **tehnici statistice non-parametrice**, mai puțin sofisticate. În schimb, variabilele măsurate prin scale numerice necesită, în cele mai multe cazuri, **tehnici statistice parametrice**, mai complexe. În cartea de față ne vom axa preponderent asupra ultimelor, deoarece în cercetarea psihologică ele constituie o majoritate clară.

1.2. Analiza descriptivă univariată

Lucrările amintite anterior şi cele prezentate în nota de subsol oferă o bună introducere în problema analizei statistice descriptive. De aceea, ne vom rezuma să prezentăm doar principalele concluzii ale unui asemenea demers descriptiv, insistând însă asupra coeficientului de corelație r, ca element fundamental în înțelegerea unor tehnici statistice complexe precum analiza factorială sau regresia liniară.

(1) Descrierea unui set de date necesită cel puțin doi parametri[1]

Unul vizează tendința centrală a datelor, adică găsirea unei valori reprezentative pentru setul respectiv de date, iar celălalt vizează gradul de împrăştiere a datelor, urmărind omogenitatea sau eterogenitatea unui grup. De pildă, dacă am încerca să descriem un grup doar prin prisma primului parametru (ex. ambele clase de elevi obțin o medie generală de 8,25) nu am putea afla dacă

[1] Mai multe informații despre algoritmul de calcul al parametrilor tendinței centrale şi ai împrăştierii pot fi găsite în orice manual introductiv, recomandabile fiind lucrările lui Hohn, Vârgă şi Măruşter (1999), Hohn (2000), Radu şi colab. (1993), ori Sava (2002).

clasele sunt similare, deoarece una ar putea fi formată din elevi omogeni, toți având media în jurul valorii de 8,25, în timp ce cealaltă clasă de elevi ar putea fi formată din două categorii diferite de elevi – foarte buni şi slabi – dar care împreună ar conduce la o medie egală cu cea a primului grup.

În funcție de scalele de măsurare utilizate se stabileşte şi parametrul tendinței centrale, respectiv al împrăştierii datelor care ar trebui calculat.

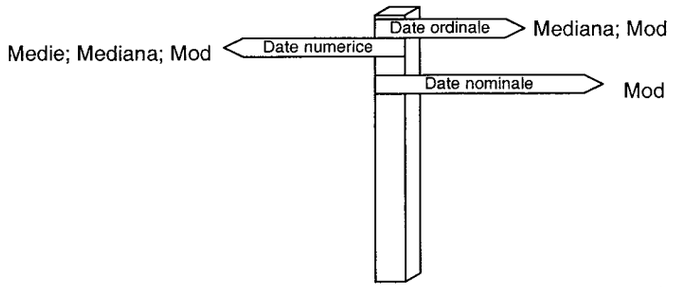


Figura 1.1. *Parametri posibili ai tendinței centrale, în funcție de tipul de variabile*

Între parametrii împrăştierii, abaterea standard şi dispersia sunt cei mai întâlniți în cazul datelor numerice, iar abaterea intercuartilă este mai des utilizată în cazul datelor măsurate prin scale ordinale. În ceea ce priveşte datele nominale, situația indicatorilor împrăştierii este neclară. Indicatori precum entropia, indicele de fragmentare, indicele de diversitate sau cel al dispersiei calitative sunt aproape necunoscuți şi lipsesc cu desăvârşire din programele statistice pe calculator. În consecință, răspândirea lor este minimă, majoritatea cercetătorilor rezumându-se la a urmări distribuția procentajelor pentru fiecare răspuns posibil (Weisberg, 1992).

În general, indicatorii împrăştierii iau valoarea zero dacă toate datele dintr-un set sunt identice. Cu cât datele sunt mai împrăştiate (diferite), cu atât valorile acestor indicatori sunt mai mari.

În cazul datelor numerice, cel mai frecvent întâlnită pereche de parametri descriptivi este media, ca indicator al tendinței centrale, şi abaterea standard, ca indicator al împrăştierii datelor. De altfel, aceşti parametri au şi cele mai multe avantaje tehnice, permițând realizarea mai multor operații precum standardizarea datelor, clasificarea răspunsurilor etc.

Totuşi, în anumite situații media şi abaterea standard pot conduce la rezultate distorsionate, iar pentru a evita acest lucru este esențial să inspectăm grafic situația rezultatelor obținute.

(2) Reprezentarea grafică însoțeşte analiza descriptivă a datelor

Reprezentarea grafică este o metodă intuitivă de a observa distribuția datelor, ea fiind cu atât mai utilă cu cât numărul de date este în creştere.

Există o multitudine de grafice posibil de realizat. Alegerea unui anumit tip de grafic se face în funcție de modul de măsurare a variabilelor, dar şi de aspecte preferențiale, ce țin de subiectivitatea fiecăruia. În funcție de tipul de date colectate, cele mai întâlnite grafice sunt:

- diagrama circulară („plăcinta") pentru variabile de tip nominal;
- histograma modificată pentru variabile de tip nominal sau ordinal;
- histograma sau poligonul frecvențelor pentru variabile numerice.

Iată un exemplu de reprezentare grafică prin metoda poligonului de frecvențe:

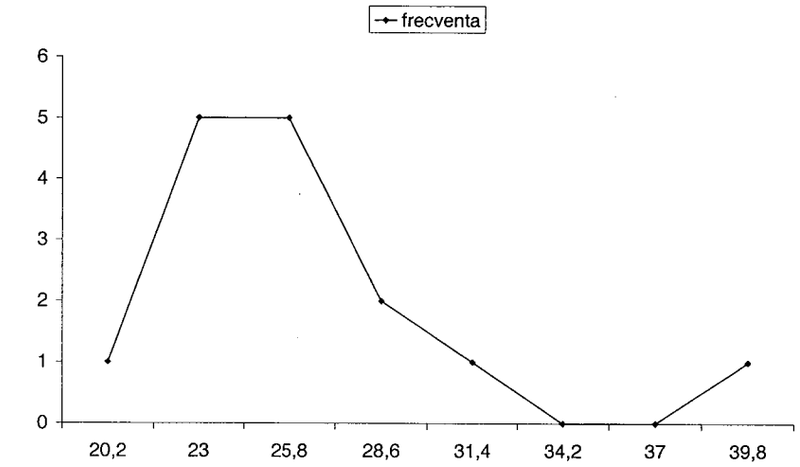


Figura 1.2. *Exemplu de poligon al frecvențelor, reprezentare tipică pentru datele numerice*

(3) Analiza datelor este influențată de forma distribuției datelor

Dincolo de aspectul pur intuitiv al reprezentării grafice, foarte importantă este interpretarea formei distribuției datelor. Aceasta ne ajută să înțelegem două aspecte esențiale: (a) care sunt parametrii tendinței centrale care reprezintă cel mai bine şirul de date?; (b) ce teste statistice pot fi utilizate pentru a testa anumite ipoteze?

Se spune că multe dintre aspectele realității au o distribuție normală. Să luăm cazul inteligenței. Cei mai mulți oameni au o inteligență medie. Întâlnim tot mai puțini oameni pe măsură ce mergem fie spre o inteligență superioară, fie spre un deficit grav de inteligență. O asemenea situație, denumită distribuție normală, poate fi reprezentată grafic sub forma unui clopot.

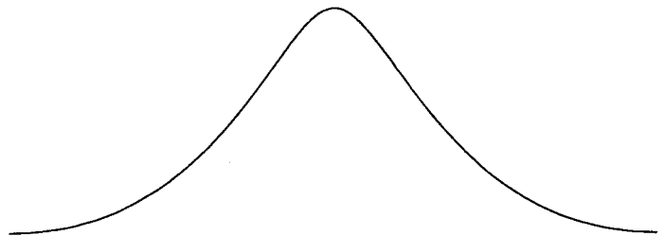


Figura 1.3. *Forma unei distribuții normale a datelor*

Multe teste statistice parametrice (utilizate în cazul datelor numerice) au drept condiție de aplicare normalitatea distribuției datelor. De aceea, este important să caracterizăm forma distribuției obținute prin intermediul a două aspecte: (a) **nivelul de boltire**, respectiv (b) **nivelul de oblicitate** (asimetrie).

(a) O distribuție normală este *mezocurtică* (normală), adică are un nivel mediu de boltire (nici prea joasă, nici prea înaltă). Totodată ea este simetrică, adică nu este înclinată evident nici spre stânga, nici spre dreapta.

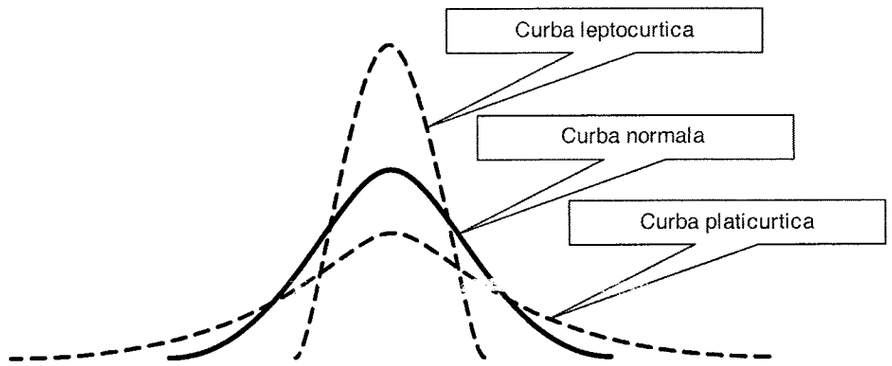


Figura 1.4. *Devieri ale unei curbe simetrice din perspectiva boltirii datelor*

Cum interpretăm distribuțiile prezentate în figura 1.4? O curbă *platicurtică* se obține atunci când subiecții au rezultate foarte împrăştiate. Nu există o frecvență mare a rezultatelor în jurul mediei, fapt ce ne conduce la ideea unui grup eterogen. Un exemplu ar fi o clasă de elevi, ale căror rezultate sunt foarte dispersate. Câțiva copii au note cuprinse între 1 şi 4, câțiva copii au rezultate între 5 şi 7, iar alți elevi au rezultate superioare, cuprinse între 8 şi 10. Reversul situației este reprezentat de o distribuție *leptocurtică*. Aceasta indică un grup omogen de date, în care cele mai multe rezultate se găsesc distribuite în jurul mediei. În cazul clasei de elevi,

aceasta ar fi alcătuită din copii cu performanțe apropiate. Cei mai mulți ar înregistra rezultate între 5 şi 7, în timp ce foarte puțini elevi ar fi incluşi în categoria elevilor cu note sub 5 sau a celor cu note peste 7. Într-o asemenea clasă lipsesc notele extreme, fapt ce face ca nu toată gama de la 1 la 10 să fie distribuită. În sfârşit, o curbă mezocurtică are un nivel mediu, normal de boltire, ceea ce înseamnă că mulți copii au rezultate apropiate de valoarea mediei, deşi există unii elevi foarte slabi, respectiv foarte buni, toată gama de note fiind acoperită.

(b) Distribuția datelor poate varia şi în funcție de nivelul de simetrie a datelor. În cazul în care distribuția este *asimetrică pozitiv,* există mai multe valori mici decât mari. Într-o situație şcolară, aceasta ar corespunde unui test dificil, în care majoritatea elevilor obțin note mici, deşi sunt şi elevi care obțin rezultate şcolare bune. În schimb, o distribuție *asimetrică negativ* indică un dezechilibru al valorilor în direcție opusă. Vor apărea mai multe rezultate mari decât mici, fapt care, transpus în note şcolare, arată predominanța notelor ridicate. În cazul probelor psihologice, prima distribuție poate fi socotită un indicator al unui grad de dificultate ridicat pentru proba respectivă, în timp ce o distribuție asimetrică negativ semnifică un grad scăzut de dificultate a probei. Bineînțeles că aceste concluzii sunt valide în condițiile în care valorile ridicate indică performanțe superioare, în timp ce valorile mici arată performanțe scăzute.

Figura 1.5. *Distribuție asimetrică pozitiv (în stânga) şi distribuție asimetrică negativ (în dreapta)*

Pentru a evita interpretările subiective, se pot calcula doi indicatori care să ajute la interpretarea formei distribuției.

Indicatorul de oblicitate (în engleză „skewness statistic") evaluează gradul de asimetrie al unei distribuții. Valoarea zero indică o distribuție perfect simetrică. Valorile negative, îndepărtate de zero, indică o distribuție asimetrică negativ, în timp ce valorile pozitive, cu cât se îndepărtează de zero, arată o asimetrie pozitivă. Pentru a diferenția o distribuție uşor asimetrică, ce poate fi tratată ca o distribuție normală, de o distribuție clar asimetrică, Lewis-Beck (1995) ia drept prag valoarea absolută de .80. Astfel, o valoare a indicatorului de oblicitate inclusă în intervalul [-.80 şi .80] poate fi tratată ca un semn al unui distribuții simetrice. Valorile mai mici de -.80 indică o asimetrie negativă, în timp ce valorile mai mari de .80 semnifică o asimetrie pozitivă. Această regulă implicită merită a fi aplicată mai ales în condițiile unui eşantion mare de subiecți, să zicem de peste 150 de persoane.

În cazul în care volumul eşantionului este mediu sau mic, putem adopta o altă strategie, bazată pe calcularea cotei z a indicelui de oblicitate. Aceasta presupune împărţirea valorii indicatorului de oblicitate la eroarea standard estimată a acestuia. Ambele elemente sunt oferite de SPSS. Field (2000) propune două praguri ale lui z în funcţie de numărul de persoane testate: 2,58 pentru un eşantion mai mic de 30 de subiecţi şi 1,96 pentru un eşantion mai mare de 30 de subiecţi. Acelaşi autor recomandă renunţarea la regulă în cazul unui număr mare de subiecţi, să zicem peste 150, deoarece z are mari şanse de a deveni semnificativ, datorită scăderii erorii standard a oblicităţii o dată cu mărirea eşantionului.

Tabelul 1.1. Exemplu de valori ale mediei, medianei şi oblicităţii privind vârsta ideală pentru căsătorie

Vârsta ideală pentru căsătorie

N	15
Media	26,00
Mediana	25,00
Oblicitatea (skewness)	2,240
Eroarea standard a oblicitatii	,580

Fiind testaţi doar 15 subiecţi, vom împărţi oblicitatea la eroarea sa standard şi vom obţine 3,86 (2,24 : 0,58). Rezultatul, în valoare absolută, este superior pragului specificat, de 2,58, prin urmare avem de a face cu o distribuţie asimetrică pozitiv. Deoarece în astfel de cazuri mediana este un parametru adecvat pentru a descrie şirul de date, am putea spune că subiecţii chestionaţi consideră vârsta de 25 de ani ca vârstă ideală pentru căsătorie. E de remarcat că operaţia trebuie completată de inspectarea graficului rezultat, deoarece distribuţiile multimodale nu pot fi detectate prin urmărirea indicatorului de oblicitate.

Nu numai oblicitatea poate fi estimată exact. Boltirea beneficiază, de asemenea, de un astfel de coeficient pentru a determina normalitatea unei distribuţii. **Indicatorul de boltire** (în engleză „kurtosis") este estimat de SPSS pe baza aceloraşi principii ca în cazul oblicităţii. Valoarea zero indică o distribuţie perfect mezocurtică, valorile pozitive, cu cât sunt mai mari, arată o tendinţă spre distribuţii leptocurtice, în timp ce valorile mai mici de zero, semnalează tendinţa spre o distribuţie platicurtică. Field (2000) recomandă calcularea lui z kurtosis după aceleaşi criterii enunţate pentru oblicitate.

Dacă există dubii cu privire la gradul de normalitate a distribuţiei datelor, se recomandă testarea condiţiei de normalitate prin testul z al lui Kolmogorov-Smirnov. Dacă valorile sunt semnificative statistic, putem afirma că distribuţia nu îndeplineşte caracteristicile unei curbe normale. În schimb, dacă z este nesemnificativ statistic, vom putea afirma că distribuţia observată nu diferă semnificativ de o distribuţie normală teoretică, deci poate fi considerată ca atare.

Din păcate, rezultatele sunt sensibile la numărul de subiecţi testaţi. Dacă obţinem mai puţin de 30 de date, există şanse mari ca z să nu fie semnificativ, cu

excepţia unor cazuri grave de depărtare faţă de o distribuţie normală. Similar, în cazul unui număr mare de date, să zicem peste 150-200, riscul de a obţine un z semnificativ este foarte mare, chiar la depărtări minore ale datelor de modelul teoretic al unei distribuţii normale.

(4) Cunoaşterea mediei şi a abaterii standard permite analize statistice complementare

În cazul unei distribuţii normale, media şi abaterea standard au o serie de proprietăţi care permit obţinerea unor informaţii descriptive sau inferenţiale complementare.

(a) *Standardizarea datelor*

De obicei colectăm datele în forma lor brută, păstrând unităţile de măsură iniţiale. În acest mod putem vorbi despre faptul că o persoană a obţinut 7 puncte la testul de agresivitate verbală, 18 puncte la un test de atenţie concentrată, 43 de răspunsuri corecte la un test de acuitate vizuală sau că măsoară 180 de cm. De cele mai multe ori nu cunoaştem ce semnificaţie au aceste cifre. Excepţie fac cazurile în care avem o idee despre evoluţia rezultatelor, precum datele antropometrice legate de înălţime, greutate ori unele informaţii cu caracter general, de tipul venitului lunar, al temperaturii etc. Spre exemplu, nu vom cunoaşte dacă rezultatul de 18 puncte la un test de atenţie concentrată este mai bun decât obţinerea a 43 de răspunsuri corecte la testul de acuitate vizuală.

Pentru a rezolva problema, se poate recurge la transformarea rezultatelor iniţiale, din cote brute, în cote standardizate. Există mai multe tipuri de transformări, toate având la bază **cotele z.**

Transformate în cote z, scorurile brute care sunt egale cu media vor avea valoarea 0. Valorile negative vor apărea pentru scoruri mai mici decât media, iar valorile pozitive, pentru rezultate superioare mediei. Mai mult, abaterea standard a datelor transformate în cote z va fi întotdeauna egală cu 1. Astfel, după transformare putem compara orice variabile, deoarece toate vor avea media 0, iar abaterea standard 1. În plus, se cunoaşte că într-o distribuţie normală, peste 99% din rezultate vor avea valori z cuprinse între -3 şi +3. Formula de calcul pentru transformarea oricărei valori în cote z este:

$$z = \frac{x - m}{s} \qquad (1.1)$$

unde X este valoarea ce urmează a fi transformată; m este media obţinută în eşantionul respectiv ori media populaţiei; iar s este abaterea standard a eşantionului sau a populaţiei (în cazul în care este cunoscută).

Astfel, dacă dorim să transformăm în cote z cele 43 de puncte obţinute la testul de acuitate vizuală, vom obţine valoarea -1,71, dacă media este 55, iar abaterea standard 7. Valoarea indică un număr de răspunsuri corecte mult mai mic decât media. Similar, dacă vom transforma în cote z rezultatul obţinut în urma

testului de atenție concentrată (m = 21; s = 3,11; iar X = 18), vom obține -0,96. Pe baza acestor transformări putem afirma că, deşi ambele rezultate sunt sub medie, performanța la testul de atenție concentrată este mai bună decât cea obținută la testul de acuitate vizuală.

Pe baza cotelor z se pot calcula şi alte cote standardizate, precum **cotele T**. Acestea apar frecvent ca modalitate de etalonare preliminară a testelor.

(b) *Clasificarea subiecților*

În cazul unei distribuții normale, media şi abaterea standard oferă mai multe informații, cum ar fi locul pe care-l ocupă un subiect într-un grup de oameni. Spre exemplu, putem determina poziția unei persoane cu un nivel de inteligență de 115 IQ, în condițiile în care inteligența medie este 100 IQ, iar abaterea standard 15 IQ. Altfel spus, prin apelul la proprietățile unei distribuții normale putem demonstra ce procent de subiecți obțin rezultate mai mari de 115 ori putem stabili câte persoane au rezultate inferioare acestei valori. Totul va deveni clar dacă vom înțelege figura 1.6.

În cazul unei distribuții normale, media împarte şirul de date în două părți egale: 50% din rezultate sunt mai mici decât media, iar 50% mai mari. În cazul de față ne aşteptăm ca 50% dintre persoane să aibă un IQ mai mic de 100, iar celelalte 50%, un IQ mai mare de 100.

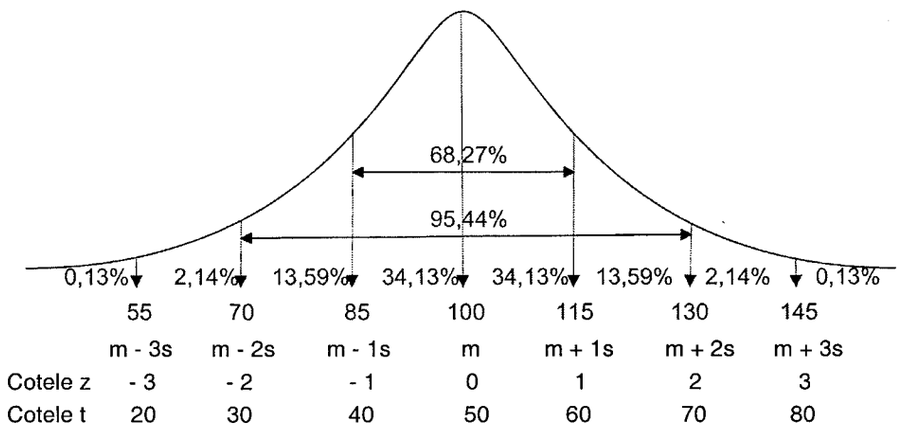


Figura 1.6. *Proprietățile unei distribuții normale*

Noi ştim că abaterea standard pentru acest eşantion este de 15 IQ. Pe baza acestei informații putem stabili procentajul de persoane cuprinse între anumite valori IQ, deoarece într-o distribuție normală întotdeauna vor fi aproximativ 34,135% dintre rezultate între medie şi valoarea aflată la o abatere standard față de medie. Revenind la exemplul nostru, estimăm că 34,135% dintre persoanele testate obțin rezultate între 100 şi 115 IQ, acelaşi procentaj având un nivel de

dezvoltare intelectuală cuprins între 85 şi 100 IQ. Dacă vom face un simplu calcul, vom constata că 68,27% dintre persoanele testate au rezultate cuprinse între 85 şi 115 IQ, adică între valorile situate la câte o abatere standard în minus, respectiv în plus, de la medie [m-1s; m+1s].

În mod similar se poate estima procentajul de persoane aflate între diferite secțiuni ale acestei curbe. Spre exemplu, procentul de răspunsuri cuprins între o abatere standard față de medie şi două abateri standard față de medie este de 13,59%. Astfel, în problema nivelului de inteligență, aproximativ 13,59% din rezultate au valori cuprinse între 70 şi 85, respectiv între 115 şi 130.

În figura 1.6 sunt prezentate şi cotele z, respectiv T, corespunzătoare procentajelor stabilite pe baza proprietății unei distribuții normale. Astfel, între valorile -2 şi -1 cote z, respectiv între 30 şi 40 cote T, se regăsesc 13,59% din rezultate. Similar, între -1 şi +1 cote z, respectiv între 40 şi 60 cote T, vom obține aproximativ 68,27%.

Pe baza acestor procentaje se pot stabili categorii calitative de includere a rezultatelor, utilizând următoarea grilă:

- *rezultate foarte slabe* (valori mai mici decât cele aflate la două abateri standard sub medie – 2,27%);
- *rezultate slabe* (valori aflate între două abateri standard sub medie şi o abatere standard sub medie – 15,39%);
- *rezultate medii* (valori aflate între o abatere standard de-o parte şi de alta a mediei – 68,27%);
- *rezultate bune* (valori aflate între o abatere standard peste medie şi două abateri standard peste medie – 15,39%);
- *rezultate foarte bune* (valori mai mari decât cele aflate la două abateri standard peste medie – 2,27%).

Proprietățile distribuției datelor observate sunt esențiale, nu doar în statistica descriptivă, ci şi în cea inferențială, cum ar fi testarea ipotezelor. Trebuie remarcat că nu toate tehnicile statistice inferențiale se bazează pe proprietățile unei distribuții normale, însă o logică similară este utilizată şi în cazul altor tipuri de distribuții. [3]

(c) *Estimarea parametrilor statistici*

Dacă analizele anterioare se plasau la nivelul lotului de subiecți testați, pe baza mediei şi abaterii standard ai eşantionului pot fi calculați şi parametrii statistici de la nivelul populației din care a fost extras eşantionul. Acest lucru se obține prin estimarea erorii standard a mediei şi a intervalului de încredere asociat acesteia.

Foarte rar poate fi testată întreaga populație, din cauza resurselor limitate, de aceea se obişnuieşte a se extrage un singur eşantion din cadrul acesteia, deşi, teoretic, pot fi extrase randomizat o mulțime de eşantioane.

[3] Cei interesați să cunoască mai multe informații despre proprietățile matematice ale diverselor tipuri de repartiții ale datelor pot consulta lucrarea lui Clocotici şi Stan (2000).

Să presupunem că vrem să măsurăm nivelul de empatie la o clasă de liceu şi obţinem un rezultat mediu de 14,3. Dacă am fi evaluat nivelul de empatie la clasa de vizavi am fi obţinut 14,7, iar dacă am fi extras un eşantion dintr-o altă clasă am fi obţinut 14,2. Continuând, în cazul extragerii aleatorii de eşantioane pentru a identifica nivelul de empatie, am obţine o serie de medii, pe baza datelor observate. Majoritatea mediilor ar lua valori apropiate de cele menţionate anterior, puţine dintre ele îndepărtându-se de acestea. Cu alte cuvinte, vom obţine o distribuţie normală a mediilor eşantioanelor extrase, care vor avea o anumită variabilitate. Gradul de împrăştiere a acestor valori se numeşte **eroarea standard a mediei**, ea fiind un estimator analog abaterii standard la nivelul unui eşantion. Aceasta poate fi calculată pornind de la datele obţinute dintr-un singur eşantion, iar formula de calcul este:

$$s_m = \frac{s}{\sqrt{n}} \qquad (1.2)$$

unde s este abaterea standard a eşantionului, iar n este volumul eşantionului. Din formulă se deduce că eroarea este mai mare în condiţiile în care datele sunt puţine şi împrăştiate, respectiv mai mică, dacă există multe date, iar grupul observat este omogen.

O variantă similară este eroarea standard a frecvenţelor, utilizată des în sondajele din sociologie, fiind cunoscută publicului larg sub numele de eroare de eşantionare. Astfel, procentul de eroare dintr-un sondaj, să zicem ± 3%, înseamnă tocmai eroarea standard a frecvenţelor.

Eroarea standard a mediei este capitală, atât pentru **testarea ipotezelor** bazate pe diferenţa dintre medii, cât şi pentru estimarea unui **interval de încredere al mediei**. Acesta include valorile între care ne aşteptăm să găsim media unei variabile la nivelul populaţiei, pornind de la media obţinută în eşantionul studiat.

1.3. Analiza descriptivă bivariată

În cazul în care două variabile sunt analizate simultan, vorbim de o statistică bivariată. Cei mai importanţi indicatori descriptivi în această direcţie sunt *indicatorii de asociere*. Asocierea sau relaţionarea dintre două variabile poate fi înţeleasă prin termenii de *covarianţă* şi *independenţă* (Neuman, 1997). Două variabile covariază dacă valorile uneia variază împreună cu valorile celeilalte. În realitate se pot constata tot felul de astfel de legături. Se observă, în general, că venitul personal ridicat covariază cu un număr crescut de kilometri parcurşi în maşina personală, că rezultatele ridicate la matematică sunt asociate cu rezultate bune la fizică, ori că un grad crescut de timiditate covariază cu interacţiuni sociale mai puţine etc. Conceptul de independenţă este opus celui de covarianţă. Atunci când valorile a două variabile nu se asociază spunem că cele două variabile sunt

independente. Un exemplu ar fi absenţa relaţiei la adulţi între ordinea la naştere şi numărul de la pantofi.

Pentru a vedea gradul de covariaţie dintre două variabile avem la dispoziţie trei elemente descriptive:

(1) calcularea coeficienţilor de corelaţie ca indicatori de asociere;
(2) reprezentarea grafică a norului de puncte (în engleză „scatterplot");
(3) realizarea de tabele de contingenţă (asociere).

1.3.1. Coeficienţii de corelaţie

Indicatorii descriptivi care arată gradul de covariaţie dintre două variabile sunt denumiţi **coeficienţi de corelaţie**. Corelaţia este o formă standardizată a covarianţei, eliminând problema măsurării datelor prin scale diferite. Astfel, dacă vom calcula covarianţa dintre venit şi numărul de ani de şcoală absolviţi, prima variabilă va contribui mai mult la valoarea covarianţei obţinute, deoarece are o amplitudine mai ridicată (venitul poate varia mai mult, să spunem între 0 lei şi 100.000.000, însă numărul de ani de şcoală absolviţi se poate afla doar între 0 şi 25, incluzând doctoratul şi câteva specializări). Mai mult, valoarea covarianţei dintre venit şi numărul de ani de şcoală va fi cu siguranţă mai mare decât cea rezultată între două variabile care sunt măsurate printr-o scală ce merge doar de la 0 la 1. Standardizarea datelor în cote z elimină aceste inconveniente, deoarece toate variabilele ajung la un nivel egal de amplitudine, cuprins aproximativ între -3,5 şi 3,5 cote z. Această standardizare va egaliza influenţa variabilelor asupra gradului de asociere dintre ele şi, mai important, va permite compararea gradului de asociere dintre două variabile cu asocierea dintre alte două variabile.

Tabelul 1.2. Alegerea coeficienţilor de corelaţie în funcţie de modul de măsurare a variabilelor

Tip de variabilă	*Dihotomică*	*Nominală* *	*Ordinală*	*Numerică*
Dihotomică	r, Φ, χ^2, r tetrachorică	χ^2, λ, C, V	τ (tau)	r, r biserial, r punct biserial
Nominală *		χ^2, λ, C, V	χ^2, λ	χ^2, λ **
Ordinală			r Spearman, τ (tau)	r Spearman, τ (tau)
Numerică				r

* Date nominale cu mai mult de două categorii

** În cazul grupării datelor numerice în clase

Există mai mulţi coeficienţi de corelaţie, selectarea lor făcându-se în funcţie de tipul de date colectate (nominale dihotomice, nominale cu mai mult de două categorii, ordinale sau numerice), dar şi de liniaritatea (monotonia) relaţiei dintre două variabile.

Toți aceşti coeficienți, cu excepția lui χ^2, sunt adimensionali, permițând comparații directe. Aşa cum remarcau Clocotici şi Stan (2000), există două scale de măsurare a gradului de asociere: cele *direcționale*, cuprinse între -1 şi 1, respectiv cele *nedirecționale*, implicând valori între 0 şi 1 (figura 1.7).

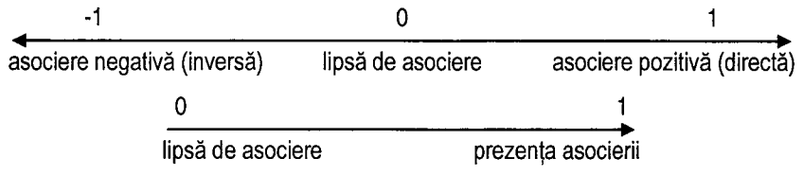


Figura 1.7. *Evoluția coeficienților de corelație direcționali (sus) şi nedirecționali (jos)*

După cum se observă în imagine, cu cât un coeficient direcțional ia o valoare mai mică de zero (minimum fiind -1), cu atât există o asociere negativă mai intensă între variabile. Prin asociere negativă înțelegem că cele două variabile covariază în sens invers: când una creşte, cealaltă scade. Un exemplu de corelație negativă este relația dintre consumul de tutun şi rezistența fizică la efort. Un consum zilnic mare de țigări tinde să ducă la scăderea rezistenței organismului la efort.

În schimb, cu cât un coeficient ia o valoare mai mare de 0 (maximum fiind 1), vorbim de o corelație pozitivă între variabile, cele două mergând în aceeaşi direcție. Spre exemplu, există o corelație pozitivă între consumul de tutun şi numărul de probleme de sănătate la nivel pulmonar. Un consum mare de tutun creşte riscul de apariție a unor boli pulmonare. În sfârşit, dacă o corelație este 0 sau foarte aproape de 0, atunci cele două variabile sunt independente, adică nu se relaționează. Un asemenea coeficient se obține, probabil, între numărul la pantofi al unei persoane şi nivelul său de inteligență, între cele două variabile neexistând o asociere.

Nu toți indicatorii iau valori cuprinse între -1 şi 1. Unii pot varia doar între 0 şi 1, unde zero semnifică absența corelației. Cu cât ne îndepărtăm de această valoare înspre 1, cu atât există o asociere mai mare între variabilele studiate.

Coeficientul de corelație r Bravais-Pearson

Datorită caracterului numeric al majorității datelor observate în psihologie, coeficientul de corelație simplă r este cel mai utilizat indicator de asociere. El este un coeficient direcțional, ceea ce înseamnă că poate lua valori în intervalul -1 şi 1. Coeficientul este cunoscut sub mai multe denumiri, precum: coeficientul de corelație liniară simplă, coeficientul de corelație a produselor momentului, coeficientul de corelație Bravais-Pearson. Dintre multitudinea de formule, cea mai simplă este:

$$r = \frac{1}{N}\sum z_x z_y \qquad (1.3)$$

unde N reprezintă numărul de subiecți testați, iar Zx şi Zy reprezintă cote standardizate ale variabilei X, respectiv Y.

Pentru a fi aplicată, ambele variabile trebuie măsurate prin scale numerice (sau în cazuri excepționale, prin variabile nominale dihotomice). În plus, relația dintre variabile trebuie să fie liniară. Pentru a înțelege mai bine ideea de liniaritate vom apela la reprezentarea grafică a norului de puncte.

1.3.2. Reprezentarea grafică a asocierii dintre variabile

Norul de puncte, cunoscut şi sub denumirea din engleză de „scatterplot", este un grafic cu două axe, fiecare dintre ele implicând una dintre cele două variabile. Nu există o regulă clară privind care dintre variabile să fie reprezentată pe abscisă şi care pe ordonată. Totuşi, după Neuman (1997), dacă putem distinge între variabila independentă (cauză) şi variabila dependentă (efect), atunci prima va fi situată pe abscisă, iar ultima pe ordonată.

X	Y	Sub.
12	5,91	1
18	7,44	2
24	8,13	3
30	8,65	4
33	9,05	5
15	8,11	6
20	7,55	7
28	8,50	8
22	7,77	9
21	6,59	10

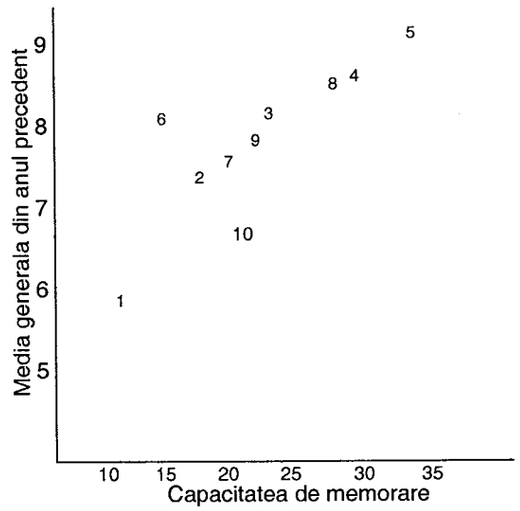


Figura 1.8. *Tabelul cu rezultate şi norul de puncte rezultat în urma reprezentării datelor*

Pentru exemplificare, vom realiza norul de puncte (diagrama de corelație) bazat pe rezultatele medii a zece subiecți la două probe: capacitatea de memorare vizuală (exprimată prin numărul de cuvinte memorate dintr-o planşă, în 30 de secunde) şi performanța şcolară (măsurată prin media generală din anul precedent). Prima va fi notată cu X, în timp ce performanța şcolară va fi notată cu

Y. Să presupunem că au fost obținute rezultatele prezentate în partea din stânga a figurii 1.8.

Am înlocuit punctele din nor cu numărul de identificare al fiecărui rezultat. Astfel, indicele 1 reprezintă locul pe grafic dat de rezultatele obținute de primul subiect testat (X = 12; Y = 5,91), indicele 2 reprezintă locul pe grafic aflat la intersecția rezultatelor celui de-al doilea subiect (X = 18; Y=7,44) ş.a.m.d.

Norul de puncte ne oferă o serie de informații importante, vizând (a) *forma* relației dintre două variabile (liniară sau non-liniară), (b) *direcția* acesteia (asociere pozitivă, negativă sau absentă), respectiv (c) *intensitatea* legăturii dintre variabile (puternică, medie, scăzută). Toate aceste aspecte vor fi discutate în continuare, nu înainte de a aminti că nivelul calitativ al interpretării bazate pe norul de puncte este dependent de numărul de subiecți testați şi, implicit, de punctele reprezentate grafic. Cu cât sunt mai multe date, cu atât precizia şi calitatea interpretării descriptive, pe baza norului de puncte, creşte.

(a) Forma relației

Relația dintre două variabile poate lua trei forme: liniară, non-liniară sau independentă, prin ultima înțelegând, de fapt, absența relației (figura 1.9).

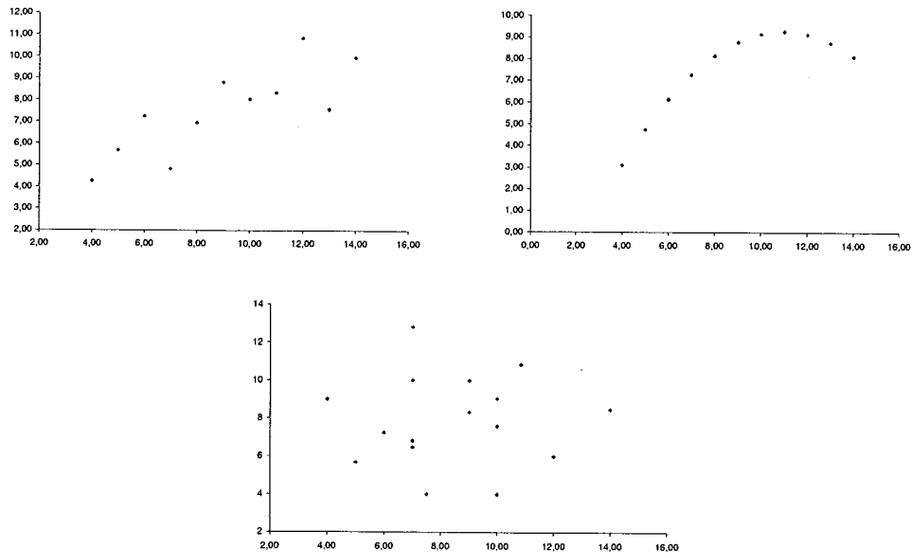


Figura 1.9. *Exemple de nor de puncte:*
relație liniară (stânga sus), relație neliniară (dreapta sus), absența relației (jos)

Dacă relația este *liniară*, atunci norul de puncte pare a avea o singură direcție, care poate fi subliniată printr-o linie dreaptă. Dacă relația este *neliniară*, norul de puncte are o direcție curbilinie, care poate fi semnalată printr-o linie curbată. În sfârşit, *absența relației* se poate observa în cadrul norului de puncte prin lipsa oricărei direcții, punctele fiind distribuite parcă la întâmplare.

Revăzând figura 1.8, între capacitatea de memorare vizuală şi performanța şcolară există o relație liniară. Se observă că norul de puncte evoluează liniar, dinspre stânga jos spre dreapta sus.

(b) Direcția relației

În cazul unei relații liniare, se poate aprecia direcția norului de puncte (figura 1.10). În aceste condiții vorbim de o *relație pozitivă*, dacă norul de puncte evoluează dinspre stânga jos spre dreapta sus. Aceasta înseamnă că valorile mici ale lui X sunt asociate cu valorile mici ale lui Y, iar valorile mari ale lui X corespund unor valori mari Y. Titulatura de direcție pozitivă vine din faptul că o astfel de distribuție a punctelor duce la o valoare pozitivă a coeficientului de corelație. O direcție pozitivă poate fi sesizată şi în graficul vizând relația dintre capacitatea de memorare şi performanțele şcolare. Prin urmare, ne aşteptăm ca valoarea coeficientului de corelație dintre cele două valori să fie pozitivă. De altfel, r rezultat a fost 0,79, confirmând estimarea făcută pe baza norului de puncte.

Relația negativă dintre două variabile poate fi sesizată grafic, dacă direcția norului de puncte evoluează dinspre stânga sus spre dreapta jos, fapt ce semnifică ideea că rezultatelor mari ale lui X le corespund rezultate mici Y şi viceversa.

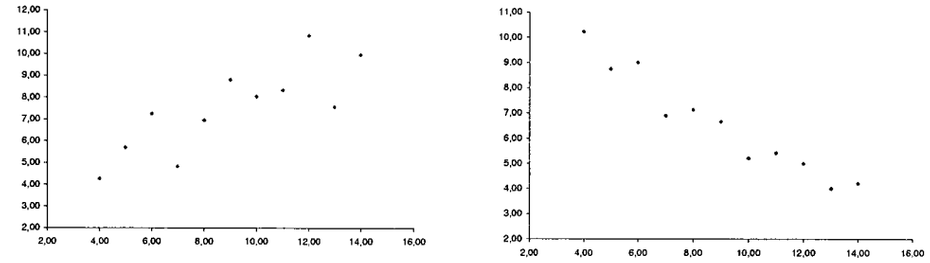


Figura 1.10. *Nor de puncte cu direcție pozitivă (în stânga)*
sau cu direcție negativă (în dreapta)

(c) Intensitatea relației

În cazul unei relații liniare, gradul de adunare a punctelor într-o anumită direcție poate varia (figura 1.11).

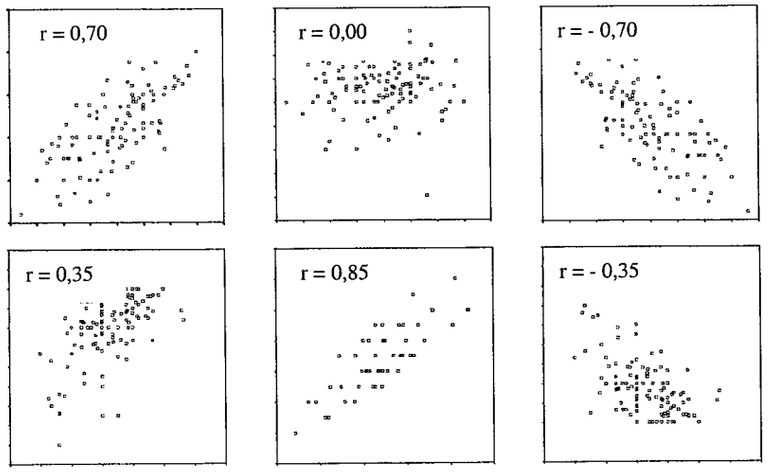


Figura 1.11. *Structura norului de puncte în funcție de nivelul relației dintre două variabile (6 exemple)*

Se observă că punctele asociate unor coeficienți de corelație mai reduşi (ex. r = .35) sunt mai împrăştiate, pe când punctele asociate unor coeficienți de corelație puternici (ex. r = .85) sunt mai apropiate de o linie imaginară, care marchează direcția de evoluție a punctelor. Cu cât punctele sunt aşezate mai apropiat de linia imaginară, cu atât intensitatea relației dintre două variabile este mai mare, fie în direcție pozitivă, apropiată de 1, fie în direcție negativă, apropiată de -1. Similar, cu cât punctele sunt mai dispersate, cu atât legătura dintre variabile este mai scăzută, ori chiar absentă, valorile coeficienților de corelație fiind mai apropiate de 0.

Importanța reprezentării grafice prin norul de puncte

Pe baza informațiilor legate de forma, direcția şi intensitatea unei relații putem obține date importante din punct de vedere calitativ. Ele ar fi fost omise, dacă ne-am fi rezumat doar la calcularea coeficientului de corelație r.

Pentru a sublinia importanța realizării diagramei de corelație (norului de puncte) în interpretarea asocierii dintre două variabile, vom prezenta nu mai puțin de patru grafice distincte, care au un singur lucru în comun: coeficientul de corelație r a fost, în toate cazurile, foarte apropiat de zero. Se ştie că un asemenea rezultat indică absența unei relații, însă vom vedea cât de înşelătoare pot fi valorile lui r, dacă vom neglija să inspectăm norul de puncte.

În continuare vom descrie cele patru situații distincte care conduc la un coeficient de corelație apropiat de 0.

(1) **Absența reală a asocierii** dintre două variabile, reprezentată printr-un nor de puncte aşezate fără nici o direcție clară, cum ar fi relația dintre nivelul de inteligență şi numărul la pantofi pe o populație de adulți (figura 1.12).

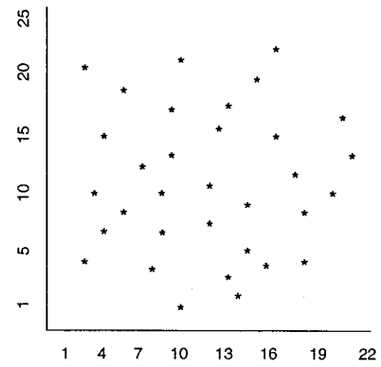


Figura 1.12. *Nor de puncte descriind absența relației dintre două variabile*

Când se obține un scatterplot ca cel prezentat în figura 1.12, valoarea 0 a coeficientului de corelație este una corectă. Cele două variabile sunt independente, deoarece modul de variație al lui X nu relaționează în nici un fel cu modul de evoluție al lui Y. Cu alte cuvinte, nu putem estima valoarea lui X când Y are o valoare mică ori mare şi viceversa.

(2) Existența a două direcții paralele ale punctelor din nor, datorate probabil unor **sub-seturi de date distincte**, cum ar fi rezultatele bărbaților față de cele ale femeilor (figura 1.13).

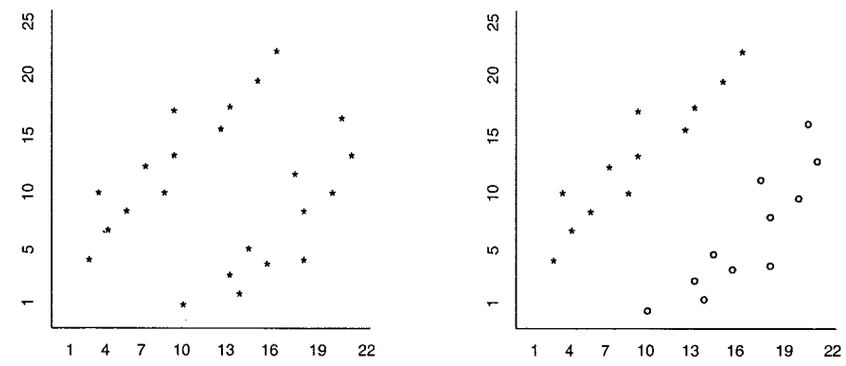


Figura 1.13. *Nor de puncte descriind o relație între două variabile mascate de prezența unei a treia (stânga – varianta inițială; dreapta – varianta în care cele două seturi sunt prezentate distinct)*

Graficul de mai sus poate fi considerat o reprezentare schematică a asociației reale dintre venitul lunar al unei persoane (abscisă) şi cheltuielile

acesteia pentru cosmetice (ordonată). În acelaşi grafic este inclusă evoluția ambelor sexe. Privind în imaginea din dreapta, cu puncte au fost reprezentate rezultatele sexului frumos, iar cu cerculeţe au fost reprezentate cele ale bărbaților. Din grafic se observă că ambele sexe, deşi evoluează paralel, au o direcție pozitivă a norului de puncte, ceea ce înseamnă că un venit ridicat covariază cu cheltuieli mai mari în domeniul cosmeticelor. Mai mult, se poate observa că bărbații, deşi au venituri salariale ceva mai ridicate, cheltuiesc mai puțini bani pe cosmetice în comparație cu femeile. Dacă nu am observa existența celor două subseturi de date şi am calcula coeficientul de corelație pentru întregul lot, am avea surpriza de a obține un coeficient de corelație apropiat de 0 şi, probabil, am concluziona greşit că nu există o legătură între venit şi cheltuielile în domeniul cosmeticelor. Greşeala s-ar datora faptului că nu am ține cont de variabila moderatoare „sexul persoanei", care influențează relația prin faptul că unele venituri ridicate ale bărbaților ar fi asociate cu cheltuieli mai reduse în comparație cu cele realizate de sexul frumos, iar unele venituri mai mici ale femeilor se vor asocia cu cheltuieli ceva mai ridicate decât ale bărbaților cu venituri superioare. Aceste aspecte vor masca evoluția liniară şi pozitivă a norului de puncte, ce indică o creştere a cheltuielilor pe cosmetice o dată cu creşterea venitului personal lunar.

(3) Existența unei **forme curbilinii**, **echilibrate**, care determină schimbarea norului de puncte dinspre pozitiv spre negativ sau viceversa, elocventă fiind relația dintre nivelul de motivare (arousal) şi performanța în sarcină (figura 1.14).

Atunci când norul de puncte are o direcție ce poate fi surprinsă doar printr-o linie curbă, putem spune că variabilele nu covariază într-un mod liniar. Această situație poate fi cauzată, în principal, de două motive: fie neglijarea unei variabile terțe (imaginea din stânga), fie existența unei relații curbilinii reale între două variabile (imaginea din dreapta). Prima situație implică prezența unei variabile moderatoare, precum în cazul discutat anterior, un alt exemplu elocvent putând fi citit în Rotariu şi colab. (1999, p.185-186). Cea de-a doua situație va fi prezentată în continuare. Dacă vom calcula coeficientul de corelație liniară r dintre nivelul de motivație în întreaga sa amplitudine (cuprinzând persoane slab motivate, optim motivate sau supramotivate) şi performanța lor în sarcină, vom avea surpriza de a contrazice una dintre cele mai vechi legi din psihologie ce afirmă că motivația (nivelul de arousal) nu afectează performanța în sarcină. Cum explicăm acest lucru? Simplu: dacă observăm imaginea din dreapta a figurii 1.14, vom constata că, până la un moment dat, cu cât nivelul de motivație este mai mare (pe abscisă), performanțele în sarcină cresc (pe ordonată). Adică subiecții foarte slab motivați obțin performanțe modeste, însă o dată cu intensificarea nivelului de arousal creşte şi performanța. Toate acestea se întâmplă până la un punct, socotit a fi un nivel optim motivațional, după care situația se inversează. Astfel, se poate constata că la intensități foarte crescute ale nivelului de arousal, performanța scade. Dacă am calcula coeficientul de corelație doar pentru subiecții aflați între un nivel slab de motivare şi un nivel optim de motivare (în stânga liniei punctate), vom obține un

rezultat pozitiv de 0,65. Vom fi tentați, astfel, să afirmăm că există o asociere pozitivă între nivelul de motivație şi performanța în sarcină, ambele evoluând în aceeaşi direcție. Similar, dacă am calcula coeficientul de corelație r doar pentru datele aflate între un nivel optim de motivare şi supramotivare (în dreapta liniei punctate), vom obține un rezultat negativ, de -0,65. Valoarea ne indică o corelație inversă între nivelul de arousal şi performanța în sarcină; cu cât gradul de motivare creşte, performanțele tind să scadă.

Pe întreg lotul de subiecți, cele două direcții opuse se anulează, fapt ce conduce la situația de a avea un coeficient de corelație egal cu zero. Această valoare ne-ar induce în eroare, deoarece spre deosebire de situația absenței legăturii dintre variabile (figura 1.12), de această dată se observă că există o covarianță în evoluția celor două variabile, doar că ea nu este liniară. De aceea, în astfel de condiții, nu este indicată calcularea coeficientului de corelație liniară r, fiind necesare alte soluții, precum cele prezentate în secțiunea 3.3.3.

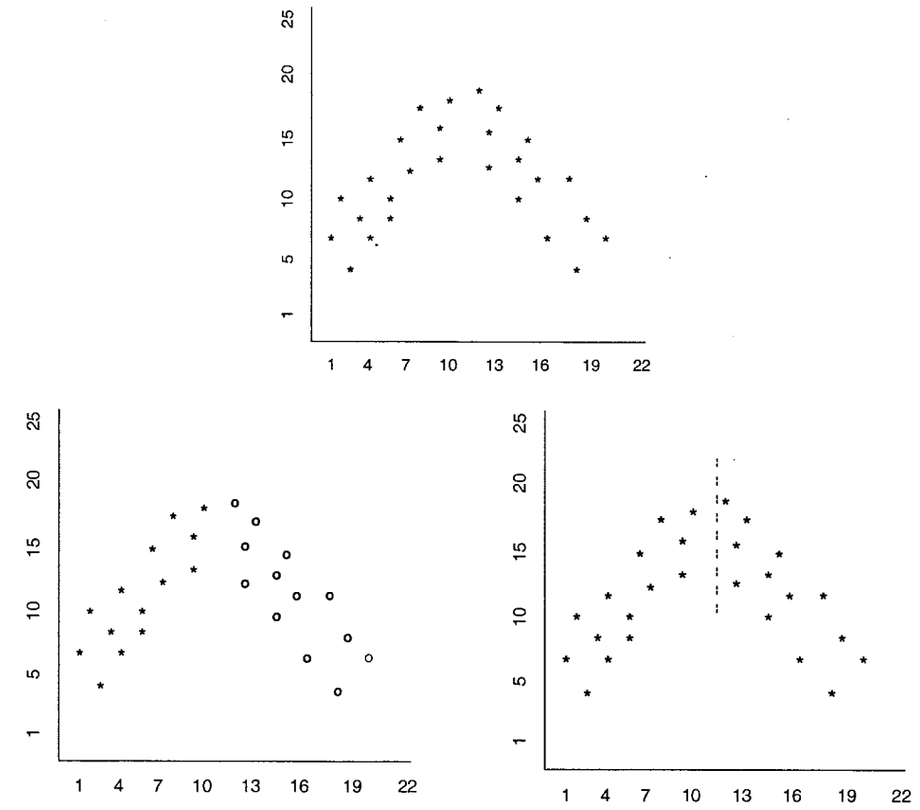


Figura 1.14. *Evoluția curbilinie a norului de puncte*
Situație datorată evoluției diferite a unor subseturi de date (stânga jos), respectiv unei relații non-liniare autentice între două variabile (dreapta jos)

(4) Existența unei asocieri clare între două variabile, mascată însă de unele **date extreme** care reduc valoarea coeficientului de corelație (figura 1.15).

O altă situație în care coeficientul de corelație este puternic distorsionat, deşi există o relație între variabile, apare în prezența unor date extreme, neobişnuite (în engleză „outliers"). Observând figura 1.15, putem distinge o relație liniară şi pozitivă, cu excepția unui singur punct, situat departe de celelalte, în dreapta jos a norului de puncte. În cazul în care datele sunt puține, ca în exemplul ilustrat, această valoare singulară poate afecta esențial valoarea coeficientului de corelație. Astfel, dacă vom calcula r pentru toate datele, exceptând valoarea respectivă, vom obține un r extrem de mare, mai exact 0,98. În schimb, dacă vom calcula r incluzând valoarea neobişnuită, r scade drastic la 0,55, iar dacă vom adăuga o altă valoare extremă, coeficientul r va scădea şi mai mult, apropiindu-se de 0.

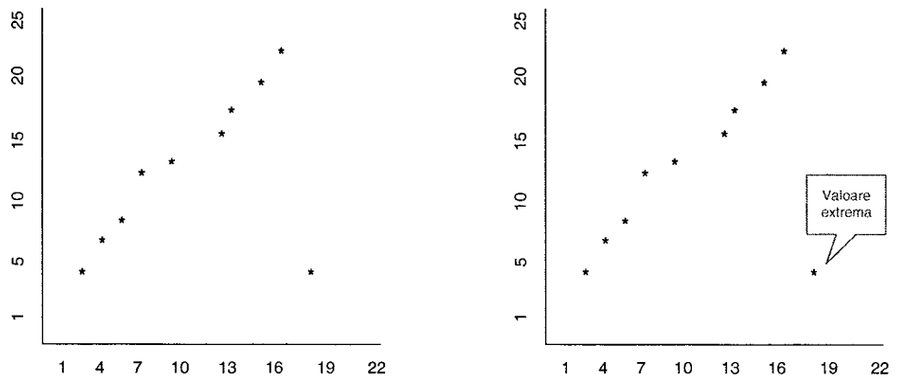


Figura 1.15. *Norul de puncte în situația existenței unor valori extreme (outliers)*

Cohen (1990) aminteşte de exemplul unui set de date a 25 de persoane, conținând informații despre înălțimea şi greutatea acestora. Coeficientul de corelație obținut a fost de 0,83, însă dacă la un singur subiect s-a inversat din greşeală ordinea de trecere a datelor, trecându-se înălțimea subiectului în coloana destinată greutății şi viceversa, r a devenit -0,26!

În astfel de situații, măsurile care trebuie luate sunt dependente de sursa apariției acestei valori neobişnuite. Dacă ea apare ca urmare a unei erori de introducere a datelor în calculator sau ca urmare a unei neînțelegeri a sarcinii ori a unei lipse de motivație din partea subiectului care a oferit acest rezultat, vom putea elimina informația din baza de date. Dacă nici unul dintre motivele de mai sus nu stă la baza apariției valorilor neobişnuite, se vor calcula doi coeficienți de corelație, rezultați prin includerea, respectiv excluderea acestor date. În fine, în cazul unui număr foarte mare de date, influența datelor extreme asupra coeficientului de corelație scade mult, iar asemenea date nu mai constituie un impediment în obținerea unor rezultate acurate.

O altă anomalie importantă, care trebuie depistată grafic, deoarece influențează nivelul de asociere dintre două variabile, se numeşte **heteroscedasticitate**. Situația dorită are o titulatură la fel de greu de reținut, **homoscedasticitate**. Din fericire, cele două situații sunt mai uşor de înțeles decât de pronunțat.

Într-o situație „sănătoasă", a unei relații liniare, norul de puncte trebuie să aibă aproximativ aceeaşi distribuție (grosime) pe întreaga amplitudine a lui X şi Y, determinând o formă omogenă denumită tehnic homoscedasticitate. Într-o distribuție „nesănătoasă" a unei relații liniare, norul de puncte nu are aceeaşi conformație pe toată distribuția lui X şi Y. Faptul este denumit heteroscedasticitate, datorită eterogenității norului de puncte. De exemplu, în imaginea din dreapta a figurii 1.16, se observă că liniaritatea datelor este mai pregnantă în prima parte a evoluției lui X, până în dreptul valorii 110, după care aspectul liniar al distribuției punctelor se păstrează într-o formă mult estompată, punctele fiind mult mai împrăştiate.

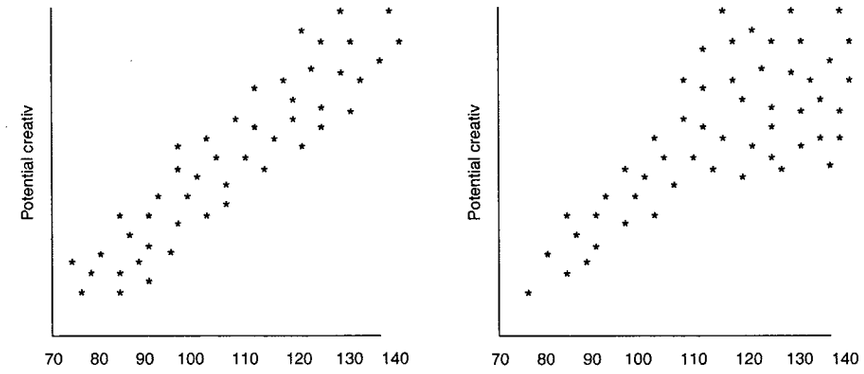


Figura 1.16. *Situațiile de homoscedasticitate (stânga) şi heteroscedasticitate (dreapta)*

Dacă ați reuşit să surprindeți ce înseamnă acest lucru din perspectiva valorii coeficienților de corelație, atunci veți intui că r este aproximativ 0,90 pentru relația dintre valorile mici ale lui X şi Y, respectiv aproape nul pentru a descrie relația dintre valorile mari ale lui X şi Y. Ironic, valoarea lui r pentru întregul set de date este 0,45, rezultat care nu reflectă nici una din cele două realități. În practică, o asemenea relație heteroscedastică se întâlneşte dacă vom dori să stabilim gradul de asociere dintre nivelul de inteligență şi potențialul creativ. Potențialul creativ este dependent, până la un anumit nivel, de inteligența unei persoane. Cu cât aceasta este mai inteligentă, cu atât creşte potențialul creativ. Dincolo de un anumit prag, situat după unii cercetători la un IQ de 110, lucrurile se schimbă, relația dintre cele două variabile diminuându-se foarte mult. Explicația pentru acest lucru este simplă: potențialul creativ are nevoie de un anumit nivel de dezvoltare

intelectuală. O dată atins acel prag, intervin alți factori, intelectuali sau non-intelectuali, care pot determina ca o persoană să se comporte mai mult sau mai puțin creativ. Astfel, importanța inteligenței în procesul creativ se estompează.

Lecția care trebuie dedusă din situația descrisă mai sus constă în importanța precizării amplitudinii variabilelor studiate. Astfel, vom obține un rezultat complet diferit, dacă vom studia relația dintre nivelul de timiditate şi frecvența discursurilor ținute public, având rezultate pe toată gama timidității (de la extrem de timid la extrem de asertiv) sau dacă vom măsura aceeaşi relație, luând doar rezultatele persoanelor foarte timide. Dacă nu vom ține seama, în ultimul caz, de caracterul selectiv al rezultatelor, vom ajunge la concluzii eronate, precum aceea că nu există o asociere între nivelul timidității unei persoane şi comportamentul său public.

1.3.3. Tabele de contingență – moduri de prezentare a asocierii dintre datele non-numerice

În cazul datelor de tip nominal sau ordinal, norul de puncte nu este potrivit pentru a descrie asocierile dintre variabile. De aceea, se folosesc **tabele de contingență** (de asociere), menite să surprindă covarianța variabilelor studiate.

Un astfel de tabel include datele de la un lot de 100 de studenți, cu privire la motivația lor de învățare (operaționalizată prin trei categorii: „foarte motivat", „parțial motivat", „nemotivat") şi la gradul lor de mulțumire față de condițiile oferite de universitatea la care studiază („mulțumit", „nehotărât", „nemulțumit").

Tabelul se realizează prin intersectarea răspunsurilor subiecților la cele două întrebări adresate. Fiecare intersecție de răspunsuri se numeşte celulă. În interiorul acesteia pot fi trecute fie frecvențele observate, fie procentajele corespondente. În plus, la marginea liniilor şi a coloanelor se pot calcula sume marginale în cazul în care se trec frecvențele în tabel, respectiv mediile marginale, în cazul în care se lucrează cu procente.

Tabelul 1.3. Tabel de contingență pe bază de frecvențe – exemplu

	Gradul de mulțumire			
Grad de motivare	Mulțumit	Neutru	Nemulțumit	Total marginal linii
Foarte motivat	14	11	1	26
Parțial motivat	7	19	19	45
Nemotivat	2	4	23	29
Total marginal coloane	23	34	43	TOTAL = 100

Acest tabel, sau unul similar exprimat în procente, stă la baza demersului statistic al datelor non-numerice. În cazul datelor nominale nu există nici o restricție cu privire la utilizarea unor anumiți indicatori de asociere. În schimb, în cazul datelor ordinale, tabelul de contingență joacă rolul norului de puncte. Prin intermediul lui vedem în ce măsură ne putem încrede în valorile coeficienților de

corelație calculați. Astfel, în cazul datelor ordinale, prin tabelul de contingență putem nota dacă există o relație monotonă între cele două variabile.

Conceptul de relație monotonă este analog celui de relație liniară, din cazul datelor numerice (Lewis-Beck, 1995). O relație este monotonă dacă există o direcție clară de evoluție a datelor. Direcția poate fi pozitivă, un grad de mulțumire ridicat se asociază cu o motivație de a învăța, respectiv negativă, când gradul de mulțumire ridicat este asociat cu un nivel scăzut de motivare pentru învățare. Cum ne putem da seama, dincolo de intuiție, de modul de asociere a două variabile ordinale? Simplu, asigurându-ne că ambele au trecute categoriile de răspunsuri într-un mod ordonat (de exemplu de la mare la mic) şi urmărind valorile trecute pe diagonala principală (stânga sus – dreapta jos) şi cele trecute pe diagonala secundară (stânga jos – dreapta sus).

Din tabelul 1.3 se observă că cele mai multe valori se regăsesc în diagonala principală (14; 19; 23), nu mai puțin de 56 de subiecți din 100 dând răspunsuri care indică o asociere pozitivă între variabile. În schimb, doar 22 de subiecți (2; 19; 1) se regăsesc pe diagonala secundară, corespunzătoare unei asocieri negative. Prin urmare, fără a trece la calcularea vreunui indicator statistic, intuim că un grad de mulțumire ridicat, cu privire la condițiile oferite de universitate, conduce la creşterea motivației pentru învățare a studenților.

Mai mult, din tabelul 1.3 deducem că relația respectă condiția de monotonie, deoarece direcția de asociere pozitivă observată necesită frecvențe ridicate pe diagonala principală. Pentru problema de față, acest lucru înseamnă că cei mai mulți oameni mulțumiți sunt foarte motivați, iar cei mai mulți oameni nemulțumiți sunt nemotivați. În tabelul 1.4 este prezentată o situație clară de încălcare a condiției de monotonie (similară unei relații non-liniare), pornind de la aceleaşi variabile.

Tabelul 1.4. Exemplu de asociere care încalcă condiția de monotonie

	Gradul de mulțumire		
Grad de motivare	Mulțumit	Neutru	Nemulțumit
Foarte motivat	14	4	23
Parțial motivat	7	19	19
Nemotivat	2	11	1

Se observă că relația nu are o direcție de evoluție. Inițial direcția relației era pozitivă, cei mai mulți oameni mulțumiți fiind foarte motivați, iar cei indecişi, parțial motivați. Apoi, în loc să observăm continuarea direcției în sensul prezenței unui număr mare de persoane nemulțumite ca fiind slab motivate (săgeata întreruptă), observăm o schimbare de direcție, astfel încât majoritatea persoanelor nemulțumite sunt foarte motivate. În astfel de cazuri, nu este indicată folosirea coeficienților de corelație specifici datelor ordinale, cum ar fi indicatorii τ al lui Kendall sau r al lui Spearman, deoarece vor conduce la rezultate eronate şi

interpretări distorsionate. Există totuşi o soluţie care constă în renunţarea de a trata variabilele ca măsuri ordinale, în favoarea celor nominale, urmată de aplicarea indicatorilor de asociere specifici datelor nominale.

BIBLIOGRAFIE

Clocotici, V., Stan, A. (2000). *Statistică aplicată în psihologie*. Iaşi: Editura Polirom.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.

Field, A. (2000). *Discovering statistic using SPSS for Windows*. London: Sage Publications.

Hohn, M., Vârgă, D., Măruşter, L. (1999). *Curs de statistică aplicată în ştiinţele sociale*. Vol.I. Timişoara: Tipografia Universităţii de Vest.

Hohn, M. (2000). *Elemente statistice în analiza fenomenelor psihice*. Arad: Editura „Viaţa Arădeană".

Kinnear, P.R., Gray, C.D. (2000). *SPSS for Windows made simple. Release 10.* Hove, UK: Psychology Press.

Lewis-Beck, M.S. (1995). *Data analysis. An introduction.* A Sage University Paper on Quantitative Applications in the Social Sciences, 07-103. Thousand Oaks, USA: Sage.

Neuman, W.L. (1997). *Social research methods. Qualitative and quantitative approaches* (3rd ed.). Needham, USA: Allyn and Bacon.

Novak, A. (1977). *Metode statistice în pedagogie şi psihologie*. Bucureşti: E.D.P.

Radu, I., Miclea, M., Albu, M., Nemeş, S., Moldovan, O., Szamoskozi, Ş. (1993). *Metodologie psihologică şi analiza datelor*. Cluj-Napoca: Editura Sincron.

Rotariu, T., Bădescu, G., Culic, I., Mezei, E., Mureşan, C. (1999). *Metode statistice aplicate în ştiinţele sociale*. Iaşi: Editura Polirom.

Sava, F. (2002). *Pagina de statistică socială*. Găsită la http://statisticasociala.tripod.com.

SPSS (2001). *SPSS Base 11.0. User's guide*. Chicago, USA: SPSS Inc.

Weisberg, H.F. (1992). *Central tendency and variability.* A Sage University Paper on Quantitative Applications in the Social Sciences, 07-083. Newbury Park, USA: Sage.

CAPITOLUL 2

TESTAREA IPOTEZELOR

Statistica **descriptivă**, foarte utilă altminteri, se mărgineşte să prezinte aspecte detaliate ale unor situaţii observate. Cei douăzeci de elevi au petrecut, în medie, 11 secunde pentru a citi textul dat; şoarecele de laborator a atins o frecvenţă de 30 de pedalări pe minut în condiţia în care pedalarea oferea accesul la hrană ş.a.m.d. Ştiinţa nu se rezumă însă doar la fapte şi la descrierea acestora. În urma unui demers ştiinţific se pot face predicţii, se pot descoperi mecanisme cauzale sau se pot găsi explicaţii pentru fenomenele observate. Toate aceste operaţii necesită un demers logic, bazat pe inferenţe, ce trece dincolo de realitatea imediat observabilă, supusă descrierii. De multe ori datele obţinute de cercetători se rezumă la situaţia existentă în anumite eşantioane, dar concluziile vizează populaţia din care acestea au fost extrase. Pentru a realiza acest lucru se apelează din nou la inferenţe, pe baza unui demers logic inductiv. Inferenţa statistică se bazează pe teoria probabilităţilor, permiţând desprinderea unor concluzii cu caracter probabilist. În practică, orice rezultat discutat în termeni de valori semnificative statistic la un prag de .05 sau .01 a corespuns unui demers specific statisticii **inferenţiale**.

După Dyer (1995), **testarea ipotezelor** reprezintă, alături de **estimarea parametrilor statistici**, unul dintre principalele aspecte ale **inferenţei statistice**. Testarea ipotezelor este o componentă fundamentală a activităţii ştiinţifice, fiind un demers riguros ce reuneşte trei dimensiuni esenţiale: (1) analiza datelor empirice; (2) realizarea unor inferenţe logice pe baza acestora şi (3) menţinerea unei atitudini sceptice legate de concluziile obţinute. Această stare de spirit se regăseşte foarte bine în ideile lui Einstein (apud Hastie şi Stasser, 2000, p.86) extinse la nivelul teoriilor:

> „Oamenii de ştiinţă nu sunt de invidiat. Natura, sau mai exact, experimentul este un judecător neprietenos şi inexorabil al muncii cercetătorului. Niciodată nu îi spune „Da" teoriei acestuia. În cele mai favorabile cazuri îi spune „Poate", iar în majoritatea cazurilor îi spune „Nu". Dacă experimentul sprijină o teorie, pentru aceasta înseamnă „Poate", iar dacă nu o sprijină înseamnă „Nu". Probabil că orice teorie va experimenta într-o bună zi pe „Nu"- cele mai multe, chiar imediat după conceperea lor".

Ipotezele unidirecționale şi bidirecționale

Ipoteza ştiințifică este o *predicție* care are capacitatea de a fi *operaționalizată* şi *testată* pentru a oferi un răspuns problemei studiate. Un exemplu de ipoteză ar fi: „persoanele cu un nivel de extroversiune ridicat tind să experimenteze în mai mică măsură o afectivitate negativă". Totuşi, din punct de vedere tehnic, există două categorii de ipoteze: unidirecționale şi bidirecționale. Modul de formulare a ipotezei cercetării determină categoria în care este inclusă ipoteza. Dacă se precizează direcția predicției, avem de-a face cu ipoteze unidirecționale. În această categorie intră formulări de genul: „există o corelație pozitivă"; „există o corelație negativă"; „grupul A este mai bun decât grupul B"; „grupul A este mai slab decât grupul B". Dacă direcția predicției nu este precizată, vorbim de ipoteze bidirecționale. Expresii precum „există o corelație între variabile", „există diferențe între loturi", reprezintă modalități tipice de formulare a ipotezelor bidirecționale, specifică fiind absența precizării direcției de evoluție a datelor. Alegerea tipului de ipoteză se face în funcție de informațiile avute anterior şi de presupunerile cercetătorului. Dacă există suficiente indicii pentru a permite formularea unor ipoteze unidirecționale, este de dorit să alegem o asemenea formulare, deoarece există şanse mai mari ca ipoteza să fie sprijinită. În cazul în care nu există indicii cu privire la modul de evoluție a datelor, se alege o ipoteză bidirecțională. Clocotici şi Stan (2000) utilizează terminologii uşor diferite pentru a denumi cele două categorii de ipoteze. Astfel, ei utilizează termenul de ipoteză unilaterală în locul celui de ipoteză unidirecțională, respectiv pe cel de ipoteză bilaterală, în locul celei bidirecționale. Pe parcursul cărții vom utiliza ambele terminologii.

Indiferent de modul de formulare a ipotezei de cercetare, testarea ipotezei se face prin apelul la conceptul de **ipoteză nulă** sau **ipoteză statistică**. Din păcate, sensul ipotezei nule a fost înțeles greşit şi prezentat denaturat în majoritatea cărților de specialitate. De aceea se impun câteva clarificări conceptuale, urmate de prezentarea a două concepte importante, dar mai puțin cunoscute, şi anume, mărimea efectului şi puterea statistică a unei cercetări.

2.1. Ipoteza nulă – clarificări conceptuale

În mod normal, orice student care a depăşit cursurile introductive de statistică şi metodologia cercetării, cunoaşte ideea că ipoteza cercetării poate fi sprijinită sau nu de evoluția ipotezei nule. Demersul este simplu: dacă ipoteza nulă (ipoteza statistică) este respinsă, atunci ipoteza alternativă este sprijinită de datele obținute. Dacă ipoteza nulă nu este respinsă, atunci ipoteza alternativă nu este sprijinită de datele obținute. Ipoteza propusă de cercetător este una din ipotezele alternative posibile. Prin urmare, sprijinirea ipotezei alternative semnifică sprijinirea ipotezei lansate de cercetător.

Prin testarea ipotezelor se înțelege testarea ipotezei nule, finalizată prin respingerea sau nerespingerea acesteia. O atenție deosebită trebuie acordată alegerii cuvintelor. Sintagme de genul „ipoteza nulă este acceptată sau confirmată" sau „ipoteza cercetării se confirmă" trebuie evitate, deoarece sunt eronate. Singurul lucru care poate fi obținut prin testarea ipotezelor este respingerea sau nerespingerea ipotezei nule.

2.1.1. Logica testării ipotezelor

Pentru a înțelege mai bine principiul testării ipotezelor vom preciza că ipoteza nulă (Ho) ne arată probabilitatea de a obține datele observate în condițiile în care ipoteza nulă este adevărată (Cohen, 1990; Cohen, 1994). Or noi am fi mult mai interesați de răspunsul la întrebarea: „Dat fiind aceste date, care este probabilitatea ca ipoteza nulă să fie adevărată?". Iată câteva caracteristici ale testării ipotezelor:

(a) *Se porneşte de la premisa că ipoteza nulă este adevărată*

Consecința acestui fapt constă în practica eronată de a atribui diferitelor praguri de semnificație sintagma „semnificativ statistic", în cazul unui p de .05, cea de „foarte semnificativ statistic", în cazul unui p de .01, ori cea de „extrem de semnificativ statistic", în situația unui p de .001. Toate aceste praguri arată doar probabilitatea de a obține datele culese în situația în care ipoteza nulă este adevărată şi nu probabilitatea de adevăr a ipotezei nule. Aşa cum am amintit, în testarea ipotezelor se porneşte de la premisa că ipoteza nulă este adevărată.

(b) *Ipoteza nulă nu poate fi adeverită, ci doar respinsă*

Ideea poate fi înțeleasă mai bine prin apelul la următoarea analogie:

Dacă plouă, asfaltul este umed.
Asfaltul nu este umed.
Deci, nu a plouat.

Acest raționament este corect şi similar cu:

Dacă ipoteza nulă este adevărată, probabilitatea de a avea aceste date este mare.
Probabilitatea este mică (nu este mare).
Deci, ipoteza nulă este respinsă.

Lucrurile se schimbă dacă vom schimba raționamentul:

Dacă plouă, asfaltul este umed.
Asfaltul este umed.
Deci, a plouat.

Raționamentul este incorect deoarece se produce eroarea logică a afirmării consecinței. Afirmația „dacă plouă, asfaltul este umed" nu este echivalentă

cu „asfaltul este umed, deci plouă". Asfaltul poate fi ud din diverse alte motive, cum ar fi stropirea sa de către maşinile de la salubrizare. Similar, confirmarea ipotezei nule este eronată.

Dacă ipoteza nulă este adevărată, probabilitatea de a avea aceste date este mare.
Probabilitatea este mare.
Deci, ipoteza nulă este adevărată.

Astfel, ipoteza nulă poate fi doar respinsă sau nerespinsă, însă niciodată confirmată.

(c) *Decizia de a respinge ipoteza nulă se ia pe baza unui prag de semnificație*

O altă problemă în testarea ipotezei nule este caracterul arbitrar al pragului ales pentru respingerea sau nerespingerea ipotezei nule. Cel mai frecvent prag ales este de .05, valoare stipulată inițial de Fisher, părintele statisticii moderne. Ulterior această valoare a fost fetişizată. Exagerând, am putea spune că publicarea sau nepublicarea studiilor, evoluția lucrărilor de doctorat sau licență, fericirea sau nefericirea cercetătorilor au devenit dependente de acest prag.

Există mai multe interpretări posibile ale acestui prag. Fără a intra în detalii, interpretarea corectă a unui rezultat semnificativ statistic poate fi rezumată simplu în expresia: „există o probabilitate de maximum 5% de a obține din întâmplare datele observate pornind de la premisa că ipoteza nulă este adevărată".

(d) *Ipoteza nulă nu înseamnă ipoteza absenței*

În sfârşit, un aspect mai puțin cunoscut este caracterul restrâns al sensului acordat ipotezei nule. Cei mai mulți cercetători înțeleg prin ipoteza nulă absența corelației sau lipsa diferenței dintre medii. Această situație se reflectă în majoritatea cercetărilor întreprinse. De altfel, utilizând simbolurile aferente, ipoteza nulă apare frecvent trecută ca: Ho: rAB = 0 în cazul corelațiilor sau Ho: $\mu A - \mu B = 0$ în cazul diferențelor dintre medii. Ipoteza nulă nu implică automat aceste condiții. Mai degrabă, ipoteza nulă este ipoteza care trebuie anulată, fie că este vorba de situații amintite (ex. Ho: rAB = 0), fie că este stipulată o anumită valoare a priori (ex. Ho: rAB = 0,25).

(e) *Testarea ipotezelor priveşte populațiile din care au fost extrase loturile testate*

Dacă analiza s-ar fi făcut la nivelul eşantioanelor, nu am fi avut nevoie de un test de semnificație cu valoare probabilistică, deoarece asocierile dintre variabile sau diferențele dintre grupe s-ar fi dedus direct din datele observate. În acest caz ne-am fi situat la un nivel de analiză descriptiv. Testarea ipotezelor vizează însă statistica inferențială, ce urmăreşte extinderea concluziilor de la nivelul lotului testat la nivelul populației din care a fost extras eşantionul respectiv.

2.1.2. Ipoteza nulă şi mărimea eşantionului

Mulți cercetători îşi aleg numărul de subiecți ce urmează a fi testați în funcție de resursele avute la dispoziție. Această modalitate nu este însă cea mai fericită, deoarece pragul de semnificație este sensibil la volumul eşantionului. Astfel, în cazul unui eşantion redus de persoane (ex. 10 subiecți), sunt mari şanse ca rezultatul să fie nesemnificativ statistic, deşi în realitate există un efect. În schimb, prin alegerea unui eşantion foarte mare (ex. 500 de persoane), sunt şanse mari de a obține un rezultat semnificativ statistic, deşi în realitate nu ar exista efectul testat. Pentru a ilustra ultima situație vom utiliza exemplul comic oferit de Thompson (2002).

> Un cercetător decide să compare inteligența măsurată prin scoruri IQ la 12000 de studenți din două oraşe diferite. Cei din primul oraş au obținut un scor mediu de 100,15 şi o abatere standard de 15, iar cei din al doilea oraş, au obținut o inteligență medie de 99,85 şi aceeaşi abatere standard. Aplicând testul z pentru două eşantioane independente obține un rezultat semnificativ statistic. Pe baza rezultatelor, cercetătorul propune înființarea unor şcoli speciale pentru supradotați în primul oraş, deoarece ei sunt semnificativ mai inteligenți decât cei din al doilea.

2.1.3. Erorile de tip I şi II în testarea ipotezelor

Interpretarea pragului de semnificație poate fi îmbogățită prin discutarea a două erori posibile ce pot apărea în testarea ipotezelor. Pentru a simplifica înțelegerea acestora voi apela la următorul exemplu. Un cercetător este interesat să studieze în ce măsură ascultarea muzicii clasice facilitează obținerea unor performanțe superioare în domeniul învățării unei limbi străine. Pentru a testa această ipoteză, el alege un lot de 50 de studenți voluntari, dornici de a învăța limba suedeză. Aceştia sunt împărțiți aleatoriu în două grupe. Prima, denumită experimentală, este formată din 25 de subiecți care vor învăța timp de patru săptămâni limba suedeză într-un mediu controlat, al cărui fundal muzical este asigurat de lucrările lui Bach şi Grieg. Ceilalți 25 de participanți, din grupul de control, vor învăța limba suedeză într-un mediu din care lipseşte orice fond muzical. După cele patru săptămâni, participanții din ambele grupe au fost testați privind cunoştințele lor de limbă suedeză.

Conform ipotezei de cercetare, ne aşteptăm ca subiecții care au învățat limba suedeză în condițiile unui fundal muzical să obțină rezultate mai bune la testul de limbă suedeză în comparație cu studenții din grupul de control. Simbolic formularea poate fi transpusă sub forma H1: $\mu E > \mu C$.

Pentru a testa ipoteza va trebui să apelăm la conceptul de ipoteză nulă. De cele mai multe ori ipoteza statistică are o formă opusă ipotezei de cercetare. În cazul nostru, ipoteza nulă ar putea fi: Nu există diferențe semnificative în performanțele de învățare a limbii suedeze între cele două grupe de subiecți. Cu

alte cuvinte, muzica clasică nu facilitează învățarea limbilor străine. Simbolic, formularea poate fi H0: $\mu E \leq \mu C$.[1]

În urma analizării datelor obținute, ne putem afla în una dintre următoarele patru situații:

(1) respingem ipoteza nulă, constatând diferențe semnificative între cele două grupe de subiecți, iar această decizie este corectă, deoarece în realitate muzica clasică afectează pozitiv învățarea limbilor străine (cadranul 1 din tabel);

(2) respingem ipoteza nulă, constatând diferențe semnificative între cele două grupe, iar această decizie este greșită, deoarece în realitate muzica clasică nu are nici o influență asupra învățării limbilor străine (cadranul 2 din tabel);

(3) nu respingem ipoteza nulă, constatând că nu există diferențe semnificative între cele două grupe, iar decizia este corectă, în realitate muzica clasică neinfluențând învățarea limbilor străine (cadranul 3 din tabel);

(4) nu respingem ipoteza nulă, constatând absența diferențelor semnificative între cele două grupe cu privire la performanțele de învățare a limbii străine, iar decizia este greșită, deoarece în realitate muzica clasică influențează învățarea limbilor străine (cadranul 4 din tabel).

Aceste combinații sunt reflectate în tabelul 2.1.

Tabelul 2.1. Erorile întâlnite în testarea ipotezelor

	În cercetare ipoteza nulă	
În realitate ...	A fost respinsă	Nu a fost respinsă
Există efect (H0 trebuia respinsă)	Decizie OK (1)	Eroare de tip II (4)
Nu există efect (H0 nu trebuia respinsă)	Eroare de tip I (2)	Decizie OK (3)

Dacă am face o analogie cu situația din justiție şi am considera ipoteza nulă ca fiind situația inculpaților „nevinovați până la proba contrarie", am obține tot patru situații:

(1) o persoană vinovată în realitate ar fi găsită vinovată şi de către justiție (cadranul 1);

(2) o persoană nevinovată în realitate ar fi găsită vinovată de justiție (eroare de tip I, cadranul 2);

(3) o persoană nevinovată în realitate este achitată şi de către justiție (cadranul 3);

[1] Modul de formulare a ipotezei nule în cazul testelor unidirecționale variază de la un autor la altul. Unii preferă o formulare exactă (ex. Ho: $\mu A = \mu B$), în timp ce alții preferă o formulare inexactă (ex. Ho: $\mu A \leq \mu B$). Ambele modalități sunt acceptabile şi conduc la acelaşi rezultat. Însă, ultima expresie subliniază mai bine ideea testului unidirecțional (Leventhal şi Huynh, 1996).

(4) o persoană vinovată în realitate este găsită nevinovată de justiție (eroare de tip II, cadranul 4).[2]

Aşadar, eroarea de tip I este situația în care cercetătorul decide să respingă ipoteza nulă, deşi aceasta nu trebuia respinsă, în timp ce eroarea de tip II este dată de cazul în care ipoteza nulă nu a fost respinsă, deşi trebuia să fie respinsă. Între cele două tipuri de erori există o relație de dependență. O dată ce cresc şansele de a comite eroarea de tip I, scad şansele de a comite eroarea de tip II şi invers.

Cum procedează cei mai mulți cercetători în realitate? Ei tind să țină seama doar de eroarea de tip I şi să ignore eroarea de tip II. Din perspectiva teoriei lui Neyman şi Pearson, pragul de semnificație de .05, stabilit a priori, nu reprezintă altceva decât eroarea de tip I. Acest prag, notat α, reprezintă o probabilitate de 5% de a obține datele culese în condițiile în care ipoteza nulă este adevărată. Ea înseamnă, de fapt, o probabilitate de 5% de a greşi, respingând ipoteza nulă când ea este adevărată. Un prag de .01 înseamnă doar o probabilitate de 1% de a comite eroarea de tip I, însă şanse mai multe de a comite eroarea de tip II.

Eroarea de tip II poate fi prestabilită şi este importantă mai ales din perspectiva puterii statistice a unei cercetări. Dacă eroarea de tip I este fixată convențional la valoarea de .05, eroarea de tip II, notată cu β, nu are o asemenea valoare. Totuşi, cel mai frecvent se alege o valoare de patru ori mai mare decât prima, adică .20. Transformând probabilitățile în şanse, vom constata că o eroare de tip II de .20, indică de patru ori mai puține şanse de a comite această eroare decât de a lua o decizie corectă (.20 / .80).

2.2. Mărimea efectului

Datorită numeroaselor probleme pe care le ridică pragul de semnificație în cadrul testării ipotezelor, tot mai mulți autori au remarcat necesitatea completării rezultatelor cu ceea ce se numeşte **mărimea efectului** (Cohen, 1990; Cohen, 1994; Fan, 2001; Thompson, 1999; Thompson, 2002). Dacă testul de semnificație răspunde la întrebarea „există o diferență sau asociere semnificativă?", mărimea efectului ne arată „cât de mare este această diferență sau asociere?". Dacă nu vom ține seama de ambele aspecte vom putea ajunge la concluzii diferite. Iată, spre exemplu, o situație personală, petrecută în urmă cu câțiva ani. Am avut posibilitatea de a testa un program de intervenție menit să amelioreze relația profesor-elevi (Sava, 2001). Un lot de profesori a fost testat înainte şi după derularea unui program de training ce s-a întins pe durata a două luni. Dacă ar fi să luăm în considerare doar pragul de semnificație, atunci programul s-a dovedit a fi un succes, deoarece a reuşit să reducă nivelul de ostilitate a dascălilor în percepția

[2] Unii autori amintesc şi de o eroare de tip III (foarte puțin probabilă), care apare atunci când ipoteza nulă este respinsă corect, dar în direcție greşită (ex. m A > m B, dar μ A < μ B).

elevilor, rezultatul fiind semnificativ chiar la un p< .01. De altfel, 14 din cei 17 profesori au reuşit performanţe mai bune în situaţia de post-test. Dacă vom ţine seama şi de mărimea efectului, rezultatele nu mai sunt atât de promiţătoare, deoarece mărimea efectului a fost mică (d = .29). Ce înseamnă acest lucru? Se pare că programul de intervenţie, deşi a avut un efect, nu a reuşit să schimbe foarte mult din percepţia elevilor în ce priveşte relaţia lor cu profesorii participanţi în program. Un profesor care anterior avea o relaţie negativă cu elevii nu a fost perceput în termeni favorabili nici după terminarea programului de intervenţie. Totuşi, imaginea lor s-a îmbunătăţit foarte uşor. Astfel, programul nu a avut capacitatea de a schimba radical relaţiile dintre profesorii instruiţi şi elevii acestora. Dacă am fi ignorat informaţiile oferite de mărimea efectului şi ne-am fi rezumat doar la cele oferite de testul de semnificaţie, am fi avut o imagine distorsionată în mod favorabil.

Caracterul informativ al mărimii efectului a început să nu mai treacă neobservat. Tot mai multe colective editoriale ale revistelor de specialitate cer afişarea valorilor mărimii efectului pe lângă cele ale testului de semnificaţie. Mai mult, în manualul de publicare editat de Asociaţia Psihologilor Americani se recomandă prezentarea valorilor mărimii efectului în manuscrisele trimise spre publicare (APA, 2001). Aşadar, mărimea efectului s-a impus tot mai mult ca un termen generic pentru o familie de indicatori care informează asupra magnitudinii unui efect.

2.2.1. Indicatori ai mărimii efectului

Pragul de semnificaţie statistică ne arată doar că există o probabilitate foarte mică de a obţine datele observate în cazul în care ipoteza nulă este adevărată, însă nu ne informează dacă această situaţie are o importanţă practică. Pentru a evidenţia acest lucru vom apela la calcularea mărimii efectului. Numărul mare de indicatori existenţi poate fi redus la două categorii de indici: bazaţi pe o diferenţă standardizată între medii, respectiv bazaţi pe procentul de dispersie explicată. În prima categorie întâlnim mai ales indicii d al lui Cohen, Δ (delta) al lui Glass sau g al lui Hedges. În cea de-a doua categorie avem r, r^2, η^2 (eta), ω^2 (omega) etc. Indiferent de categoria din care fac parte, indicii pot fi uşor transformaţi dintr-o categorie într-alta. Iată formulele pentru transformarea reciprocă a celor mai întâlniţi indicatori ai mărimii efectului, d şi r (Rosenthal, Rosnow şi Rubin, 2000):

$$d = \frac{2r}{\sqrt{1-r^2}} \qquad (2.1) \qquad r = \sqrt{\frac{d^2}{d^2+4}} \qquad (2.2)$$

Vom prezenta pe parcursul fiecărui capitol indicatorii adecvaţi ai mărimii efectului. În consecinţă, în continuare nu vom insista asupra particularităţilor fiecărui indicator amintit.

2.2.2. Interpretarea mărimii efectului

Magnitudinea unui efect poate fi interpretată în mai multe moduri. Utilizând criteriile iniţiale propuse de Cohen putem distinge trei nivele de intensitate: efect minim sau efect scăzut; efect mediu şi efect puternic sau mare. Valorile corespondente aproximative pentru fiecare tip de efect în cazul lui d şi r^2 sunt prezentate în tabelul 2.2.

Conform acestor valori, se observă că mărimea efectului obţinută în urma programului de intervenţie vizând ameliorarea relaţiei profesor-elev este una redusă (d = .29, iar r^2 = .02). Aceste valori pot fi interpretate şi altfel, în funcţie de tipul de indicatori ai mărimii efectului folosiţi. Astfel, dacă s-a utilizat coeficientul de determinare r^2, vom putea spune că doar 2% din dispersia datelor reflectând relaţia profesor-elev a fost influenţată de programul de intervenţie aplicat. În schimb, dacă vom apela la calcularea lui d, vom deduce că în urma aplicării programului de intervenţie media grupului a ajuns la centilul 62 în comparaţie cu centilul 50, corespunzător mediei grupului înainte de începerea intervenţiei. Centilele sunt utilizate în ierarhizarea rezultatelor, spre exemplu valoarea 50 însemnă un rezultat mediu, iar centilul 90, un rezultat foarte bun, caz în care doar 10% din rezultate sunt mai bune, în timp ce 90% sunt mai slabe. O mărime a efectului d de .00 corespunde centilului 50, în timp ce un d de .20 arată centilul 58, un d de .50 corespunde centilului 69, iar d de .80 indică centilul 79. Mai multe amănunte pentru cei interesaţi pot fi întâlnite în Becker (2003).

Tabelul 2.2. Valorile reper d şi r^2 pentru mărimea efectului

Mărimea efectului	d al lui Cohen	coeficientul de determinare r^2
Efect scăzut	în jurul a 0.20	.01
Efect mediu	în jurul a 0.50	.06
Efect puternic	în jurul a 0.80	.14

Interpretarea rezultatelor obţinute într-o cercetare trebuie făcută însă atât în funcţie de mărimea efectului, cât şi în funcţie de pragul de semnificaţie obţinut. Fan (2001) oferă o bună sinteză a interpretării tuturor posibilităţilor întâlnite în practică. O adaptare a concluziilor sale este prezentată în tabelul 2.3.

Revenind la programul de intervenţie menit să amelioreze relaţia profesor-elev, observăm, din tabelul 2.3., două explicaţii posibile pentru situaţia unui test semni-ficativ statistic (p < .05) şi a unui efect de intensitate scăzută. Fie există mari şanse de a se fi comis eroarea de tip I, fie rezultatele, deşi semnificative, nu au o importanţă practică. Prima explicaţie nu pare verosimilă, ţinând cont de numărul

mic de subiecți utilizați şi de puterea statistică redusă. Prin urmare, cel mai probabil, ne aflăm în situa-ția în care programul contribuie în mică măsură la ameliorarea relației profesor-elev.

Tabelul 2.3. Ghid simplificat de interpretare a mărimii efectului şi a testului de semnificație

Mărimea efectului	Testarea ipotezelor	
	Ipoteza nulă nu este respinsă	*Ipoteza nulă este respinsă*
Mică	1. Se pare că ipoteza nulă este sprijinită, nu există nici un efect semnificativ statistic sau practic.	1. Deşi semnificative statistic, rezultatele nu au o importanță practică deosebită. 2. Există un risc crescut de a comite eroarea de tip I (efectul detectat de fapt nu există). Ea apare datorită unei puteri statistice foarte mari.
Medie	1. Mărimea efectului arată promițător, dar există riscul ca efectul să fi apărut din întâmplare. 2. Există un risc crescut de a comite eroarea de tip II (există un efect, dar nu a fost detectat). Probabil că puterea statistică de a detecta efectul este mică.	1. Efectul obținut este puțin probabil să fi apărut din întâmplare. 2. Efectul pare a fi destul de important din punct de vedere practic. În plus, efectul este semnificativ statistic.
Mare	1. Fie efectul a apărut din întâmplare, fie rezultatul nesemnificativ statistic se datorează numărului redus de subiecți. Creşteți puterea cercetării. 2. Risc crescut de a comite eroarea de tip II.	1. Concluzionați cu încredere că efectul este semnificativ atât statistic, cât şi practic.

În sfârşit, trebuie menționată convenția legată de semnele asociate indicatorilor mărimii statistice. Acestora li se atribuie un semn pozitiv în cazul în care efectele observate sunt în direcția anticipată. În caz contrar, indicatorii mărimii efectului vor avea un semn negativ. Informația este importantă mai ales în realizarea de meta-analize. De exemplu, dacă o ipoteză anticipă rezultate superioare ale grupului A în fața grupului B, iar datele sprijină acest lucru, mărimea efectului va avea un semn pozitiv. În schimb, dacă rezultatele obținute indică o situație opusă celei exprimate în ipoteză, precum în situația în care grupa B obține rezultate superioare grupei A, rezultatului mărimii efectului i se va asocia un semn negativ (ex. $r^2 = -.17$).

2.2.3. Stabilirea mărimii efectului

Pentru a realiza un bun design al cercetării este necesar să stabilim a priori valoarea aşteptată a mărimii efectului. Rațiunea pentru care ar trebui să

facem acest lucru constă în selectarea numărului de subiecți necesari pentru a fi testați în vederea surprinderii efectului prezis. Pentru aceasta avem la dispoziție mai multe strategii de lucru (Singer, A.D. Lovie şi P. Lovie, 1986).

Prima alternativă implică valorificarea rezultatelor obținute în studii anterioare. Spre exemplu, în studierea efectului pe care muzica clasică îl are asupra învățării unei limbi străine ar putea fi consultate eventualele cercetări anterioare în domeniu pentru a putea face predicții asupra nivelului de influențare exercitat de prima variabilă asupra celei de-a doua.

A doua strategie, oarecum similară, constă în realizarea unui studiu pilot, incluzând doar câțiva subiecți, pentru a estima cât mai bine valorile mediei, abaterii standard sau a altor indicatori care pot prezice valoarea mărimii efectului.

În sfârşit, cea de-a treia strategie porneşte de la premise teoretice, utilizând valori a priori alese pe baza convențiilor propuse de Cohen. Din această perspectivă, în lipsa unor informații din literatura de specialitate, cei mai mulți cercetători aleg un nivel mediu al efectului (d = .50) şi în funcție de acesta îşi stabilesc numărul de subiecți necesari a fi testați.

2.2.4. Limitele mărimii efectului

Vom afirma, fără îndoială, că utilizarea indicatorilor de evaluare a mărimii efectului îmbogățeşte informația statistică. Mai mult, indicatorii au devenit indispensabili în realizarea meta-analizelor cantitative. Cu toate acestea, nici mărimea efectului nu este lipsită de unele neajunsuri. Dintre cele mai semnificative amintim o anumită dependență de mărimea eşantionului, respectiv caracterul relativ al importanței practice a unui efect.

Deşi s-a evidențiat că testul de semnificație poate fi redus la produsul dintre mărimea efectului şi mărimea eşantionului (Rosenthal şi DiMatteo, 2001), nici mărimea efectului nu este scutită de o anumită dependență de mărimea eşantionului. Cei mai afectați indici sunt cei bazați pe gradul de dispersie explicată, de genul r^2. Acesta reprezintă pătratul coeficienților de corelație liniară de tip r. Se cunoaşte că un număr mic de subiecți conduce la un r ridicat, în timp ce un număr foarte mare de subiecți duce la un r scăzut.

Cealaltă limită vizează relativitatea stabilirii importanței practice a unui efect. Deşi criteriile propuse de Cohen îşi păstrează validitatea în multe cazuri, există situații în care lucrurile se schimbă. Spre exemplu, mărimea efectului datorată ingerării zilnice a unei doze mici de aspirină asupra prevenirii infarctului era $r^2 = 0{,}001$, în condițiile în care r era doar 0,034. Acest rezultat infim arată că doar 0,1% din dispersia infarctului poate fi explicată de ingerarea de aspirină în doze mici. Totuşi, realizând transformările necesare şi luând în considerare valoarea umană, un asemenea rezultat ar duce la prevenirea eficientă a infarctului pentru 34 de persoane la 1000 de locuitori, fapt care nu este de neglijat (Rosenthal şi DiMatteo, 2001). Similar, în studiile vizând calitatea vieții pentru persoane cu diferite deficiențe, orice ameliorare poate avea o valoare practică, chiar dacă la

nivelul schimbărilor, acestea nu sunt unele radicale. Considerentele acestea l-au făcut pe Thompson (2002) să afirme că există trei înțelesuri ale cuvântului semnificativ. Astfel, rezultatele pot fi semnificative statistic ($p < .05$), semnificative în plan practic, din punct de vedere al magnitudinii schimbării (d sau r^2 cât mai mari), respectiv semnificative din perspectivă clinică sau valorică, în funcție de importanța domeniului vizat (cazul aspirinei sau cel al ameliorării calității vieții persoanelor cu deficiențe diverse).

2.3. Puterea statistică a unei cercetări

Imaginați-vă puterea statistică drept o capacitate de a detecta vizual fenomene. Dacă puterea statistică este mică, va fi dificil să observăm fenomenele de interes şi vom concluziona, poate, că fenomenele respective nu există. Acest lucru s-ar întâmpla datorită capacității noastre reduse de detectare vizuală a fenomenului în cauză. În schimb, dacă vom reuşi să ne îmbunătățim aparatele de detectare a fenomenului, adică să creştem puterea statistică a unui cercetări, vom avea mai multe şanse să detectăm fenomenele aşteptate. Mai mult, în situații în care puterea statistică a devenit foarte mare, vom putea detecta orice fenomene, inclusiv pe acelea minuscule sau insignifiante. În practică se întâmplă mai rar acest lucru, deoarece majoritatea studiilor suferă de o putere statistică scăzută. De altfel, Tversky şi Kahneman (1993) sau Smith (2000) reamintesc de un studiu mai vechi al lui Cohen, care identifica, într-o statistică incluzând studiile publicate în diverse reviste de specialitate, o valoare mediană a puterii statistice de .46. Acest rezultat este mult sub valoarea optimă prescrisă de .80. Situația nu s-a schimbat de atunci, rezultate modeste, similare, fiind obținute şi după 25 de ani de la efectuarea acelui studiu (Cohen, 1990).

Puterea statistică poate fi determinată foarte uşor, dacă ținem seama de valoarea lui β (riscul de a comite eroarea de tip II), deoarece puterea statistică este egală cu 1 – β. Astfel, dacă vom alege un β de .20, puterea statistică a cercetării respective va fi de .80, ceea ce semnifică o probabilitate de 80% de a detecta un efect dacă acesta există. Similar, un β de .05, conduce la o probabilitate de 95% de a identifica un efect, oricât de mic ar fi el, cu condiția ca el să existe. Importanța practică a acestor informații este capitală în construcția designului de cercetare, deoarece puterea statistică a unei cercetări poate fi modificată în funcție de trei parametri importanți: *mărimea efectului* („Cât de mare trebuie să fie un efect sau o relație dintre două variabile pentru a fi identificată?"); *mărimea eşantionului* („De câți subiecți avem nevoie pentru a observa fenomenul de interes?"); *valoarea* α („Care este pragul de semnificație de la care vom respinge ipoteza statistică?"). Pentru calcularea puterii statistice există programe precum Gpower (Faul şi Erdfelder, 1992), ce poate fi descărcat gratuit la adresa http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/index.html.

De asemenea, inversând demersul logic, putem calcula numărul de subiecți de care avem nevoie pentru a realiza o cercetare cu o putere statistică suficient de mare pentru a reduce şansele de a comite eroarea de tip II. Vă reamintim că mediana puterii cercetărilor publicate în anumite reviste de specialitate nu a depăşit .50, ceea ce înseamnă că erau şanse egale (.50 / .50) de a observa un efect atunci când acesta exista sau de a comite eroarea de tip II. Iată două exemple concrete de utilizare a informațiilor prezentate mai sus, pentru a înțelege importanța utilizării conceptului de putere statistică a unei cercetări.

În primul caz, vom presupune că un cercetător s-a arătat interesat de a studia în ce măsură există o corelație pozitivă între două probe de inteligență, Matricele Progresive Raven – varianta standard şi scorul general obținut la WAIS. Să presupunem că el a testat 25 de persoane şi a obținut o corelație r de .30. Această valoare nu este semnificativă statistic, deoarece p obținut a fost de .07, valoare superioară pragului maxim acceptat de .05. În aceste condiții, cercetătorul nu va putea respinge ipoteza nulă şi va afirma că cele două teste de inteligență nu corelează, după care, probabil, se va lansa în tot felul de speculații legate de domeniile distincte pe care cele două probe le măsoară. Totuşi, aceste teste ar fi trebuit să coreleze într-o anumită măsură, deoarece, în principiu, fac referire la acelaşi construct. În astfel de cazuri ne putem pune întrebarea: „chiar nu există o corelație între cele două probe sau rezultatul este o consecință a riscului ridicat de a comite o eroare de tip II?", adică de a ne înşela şi de a nu observa corelația dintre cele două teste. Pentru a răspunde la întrebare vom putea face un simplu calcul apelând la Gpower. Au fost testați 25 de subiecți, valoarea α prestabilită a fost de .05, ipoteza era de tip unidirecțional (există corelație pozitivă?), iar mărimea efectului a fost r^2 de .09. Rezultatul obținut indică o putere a cercetării de .45, ceea ce înseamnă mai puține şanse de a observa corelația între cele două variabile, dacă ea există şi are o valoare r de aproximativ .30, decât de a comite eroarea de tip II. În aceste condiții, concluziile legate de nerespingerea ipotezei nule şi de afirmarea absenței corelației dintre cele două probe devin extrem de discutabile şi atacabile. Dacă cercetătorul ar fi cunoscut importanța puterii statistice, ar fi putut planifica mai bine studiul, pentru a ajunge la o valoare acceptabilă a riscului de a comite eroarea de tip II. Astfel, dacă dorea să testeze existența unei asocieri pozitive între cele două variabile, în condițiile în care se aştepta la o mărime a efectului în jurul lui .09 ($r = .30$), se mulțumea cu un $p < .05$, iar puterea cercetării dorite era de .80 ($\beta = .20$), ar fi avut nevoie de 64 de subiecți pentru a testa relația dintre probe. Dacă ar fi procedat astfel, şansele de a comite eroarea de tip II ar fi fost mult reduse, facilitând astfel interpretarea rezultatelor.

Cel de-al doilea exemplu ar putea servi drept model în realizarea unui design adecvat din perspectiva erorilor de testare a ipotezelor statistice. Să presupunem că un cercetător doreşte să studieze rolul feedback-ului în rezolvarea unor sarcini noi. În acest scop, el grupează aleator participanții la studiu în patru grupe: unul de control, unul alcătuit din subiecți care primesc doar feedback negativ legat de sarcină; un al treilea grup care primeşte doar feedback pozitiv

legat de sarcină, respectiv un al patrulea grup care primeşte atât feedback pozitiv, cât şi negativ. Variabila dependentă este măsurată în secunde, vizând rapiditatea rezolvării sarcinii de către subiecți. Subiecții din cele trei grupe experimentale primesc feedbackul în timpul rezolvării sarcinii, în timp ce subiecții din grupul de control nu primesc nici un fel de feedback. Într-un astfel de caz, cercetătorul poate planifica acele caracteristici ale designului, astfel încât să aibă erori de tip I şi II cât mai reduse.

Probabil cercetătorul va testa ipoteza apelând la ANOVA simplă, urmată în caz de succes de o serie de comparații post hoc. În aceste condiții, el poate face o serie de calcule pentru a determina volumul eşantionului în condițiile în care sunt patru grupe, α dorit este de .05, puterea statistică aşteptată este de .80, iar efectul feedback-ului asupra duratei de rezolvare a sarcinii este unul mediu (f = .25). În urma calculului a rezultat că cercetătorul va avea nevoie de un eşantion de 180 de persoane, adică de 45 de subiecți în fiecare grupă pentru a putea testa optim ipoteza şi a menține scăzute erorile. În cazul în care ar fi ales 100 de subiecți, câte 25 în fiecare grupă, ar fi obținut o putere a cercetării de .50, ceea ce ar fi însemnat şanse egale de a observa efectul feedbackului asupra performanței, în cazul în care acest efect există şi are o intensitate medie sau de a comite eroarea de tip II. Mărimea efectului poate fi stabilită, fie pe baza rezultatelor obținute din cercetări anterioare, fie pe bază teoretică, pornind de la convențiile discutate anterior care împart mărimea efectului în: efect mic (influență redusă a variabilei independente (VI) asupra variabilei dependente (VD)); efect mediu (influență moderată a VI asupra VD) şi efect mare (influență puternică a VI asupra VD).

Douăsprezece modalități de a creşte puterea statistică într-o cercetare

Ținând cont că majoritatea cercetărilor suferă de o putere statistică redusă, se impun o serie de măsuri menite să îmbunătățească această situație. Deşi extrem de importantă, secțiunea poate fi sărită de către cititorii mai puțin familiarizați cu conceptele statistice, datorită dificultății crescute de înțelegere a textului în lipsa unor cunoştințe statistice şi metodologice de nivel intermediar. Totuşi, o dată dobândite aceste informații, cititorii vor putea parcurge măsurile stipulate, lectura fiind deosebit de utilă din punct de vedere practic.

Aceste măsuri pot fi împărțite în trei categorii mari: orientate spre designul cercetării (1-7), orientate spre măsurările efectuate în cadrul cercetării (8-10) şi măsuri exclusiv statistice (11-12). Aceste modalități se bazează pe lucrările lui Embertson şi Reise (2000), Hinkle, Wiersma şi Jurs (1994), Kirk (1982), Pedhazur şi Schmelkin (1991), Smith (2000) sau West, Biesanz şi Pitts (2000), cu toții interesați să continue munca depusă de Jacob Cohen, părintele puterii statistice. Majoritatea acestor măsuri au fost prezentate în altă parte (Sava, 2002; Sava, 2003).

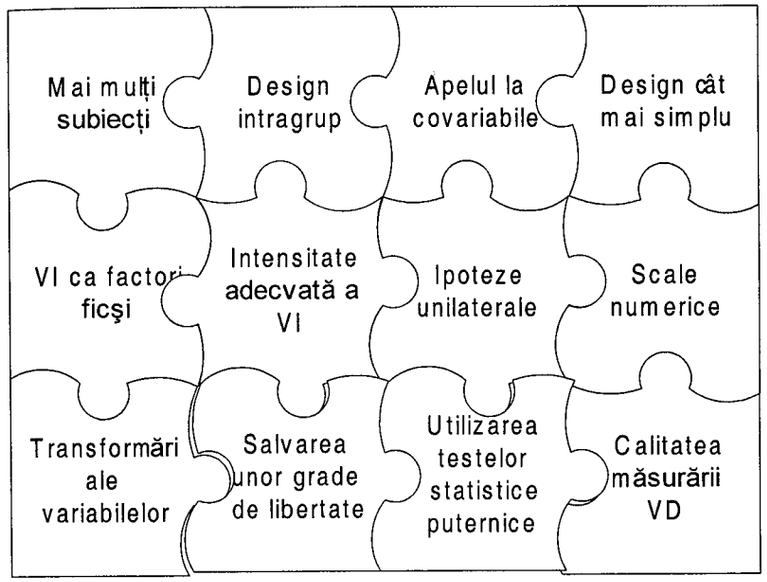


Figura 2.1. *Surse utile pentru a creşte puterea statistică a unei cercetări*

1. *Creşterea numărului de subiecți testați*

Cea mai simplă şi comună metodă de a creşte puterea statistică a unei cercetări constă în includerea mai multor participanți în studiu. Aceasta poate fi făcută fie prin creşterea numărului de subiecți din fiecare lot testat, fie prin mărirea numărului de subiecți din anumite grupe, mai importante din perspectiva ipotezelor testate. Spre exemplu, avem un design unifactorial intergrup în care, prin intermediul a trei grupe de subiecți (T1-grup de control; T2- terapie individuală; T3 – terapie de grup), dorim să testăm efectul terapiei cognitiv-comportamentale în reducerea gradului de timiditate. În condițiile în care fiecare grup este format din 10 subiecți, avem la dispoziție două posibilități de a creşte puterea cercetării prin mărirea volumului eşantionului. Prima constă în dublarea numărului de subiecți din fiecare grupă, ajungând la 60 de subiecți testați. Metoda merită utilizată când toate grupele au o importanță aproximativ egală. Dacă grupele au o importanță diferită, cum ar fi situația în care dorim să aflăm dacă intervenția de orice fel (individuală sau în grup) este semnificativ mai eficientă decât non-intervenția, vom apela la o strategie diferită. De exemplu, vom adăuga doar 5 subiecți în grupele experimentale, respectiv 20 de subiecți în grupul de control. Similar, dacă este mai important să determinăm care dintre formele de intervenție (individuală vs. în grup) este mai eficientă, vom putea creşte grupul de control cu 6 subiecți, în timp ce în fiecare grup experimental vom adăuga 12 subiecți. Se observă că numărul de subiecți adăugați se menține constant la valoarea de 30, însă strategia este diferită în funcție de ipotezele urmărite.

2. *Utilizarea unui design intragrup (cu măsurători repetate)*

În cazul în care avem acelaşi număr de subiecţi, un design intragrup are o putere statistică mai mare decât designul de tip intergrup, datorită scăderii dispersiei erorii (neexplicate) în primul caz. Fiind vorba de aceiaşi subiecţi testaţi în situaţii diferite şi nu de subiecţi diferiţi testaţi în condiţii diferite, vom avea o variabilitate mai redusă a rezultatelor. Aproape orice test statistic are, la numitor dispersia erorii şi în consecinţă, va duce la un scor t sau F mai mare, fapt ce măreşte şansele de a obţine un rezultat semnificativ statistic.

3. *Introducerea unei covariabile eficiente*

În situaţiile în care nu este posibilă creşterea volumului eşantionului din cauza economiei de timp şi de resurse sau în situaţiile în care designul de tip intragrup nu este potrivit, se poate ţine seama de efectul unei terţe variabile, denumită *covariabilă*. Spre exemplu, cunoscând importanţa încrederii în terapeut, un cercetător poate măsura şi controla influenţa acestei variabile atunci când testează efectul intervenţiei cognitiv-comportamentale în reducerea timidităţii. Prin introducerea unei covariabile, care influenţează relaţia dintre alte variabile, fie ele VI şi VD, de cele mai multe ori se reduce dispersia erorii (neexplicată), şi astfel cresc şansele obţinerii unui rezultat semnificativ statistic, fără a recurge la creşterea numărului de subiecţi testaţi. Acest raţionament poate fi vizualizat în figura 2.2.

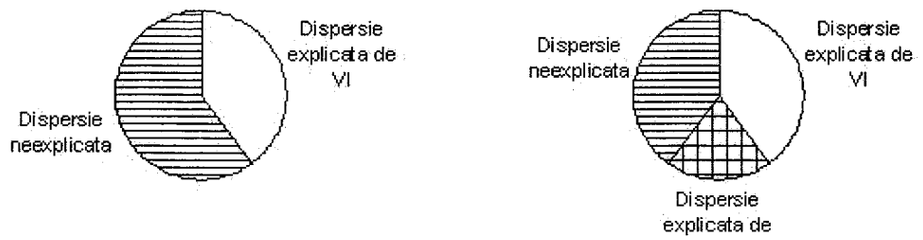


Figura 2.2. *Sursa dispersiei în ANOVA (stânga), respectiv ANCOVA (dreapta)*

4. *Utilizarea unui design de cercetare cât mai simplu*

Acest principiu filosofic, cunoscut sub numele de briciul lui Ockham, exprimă primatul unor explicaţii simple şi corecte în dauna unor modele complicate şi dense, cu prea multe variabile. În statistică, acest principiu a fost etichetat de Cohen (1990) sub forma „mai puţin este mai bine". În studii cu foarte multe variabile creşte şansa de a comite eroarea de tip I, adică de a obţine anumite rezultate semnificative din întâmplare. Vom ajunge în cealaltă extremă, întrebându-

ne care dintre rezultatele semnificative obţinute sunt reale? Prin aplicarea unor corecţii vom putea reduce distorsiunile create, dar vom creşte şansele de a comite o eroare de tip II. De exemplu, prin corecţia Bonferroni, aplicată la o matrice de corelaţii între 16 variabile (de exemplu, factorii primari de la 16PF ai lui Cattel), pragul de semnificaţie pentru respingerea ipotezei nule se va modifica de la .05 la .0004 (.05/60). Astfel, o corelaţie al cărei prag de semnificaţie obţinut este de .001 nu va putea respinge ipoteza nulă. Corecţia este absolut necesară, deoarece în absenţa ei, în condiţiile în care *nu* există asocieri reale între variabile, probabilitatea de a obţine, din întâmplare, unul sau mai multe rezultate semnificative este de 99,3%.

5. *Creşterea numărului de trepte pentru VI sau / şi creşterea amplitudinii factorilor*

Altfel spus, sunt mai mari şansele de a obţine un rezultat semnificativ statistic în condiţiile în care factorul-cauză variază mult. După trei şedinţe de terapie am putea concluziona că intervenţia nu a condus la ameliorarea situaţiei clientului. După douăsprezece şedinţe există şanse mai mari de evoluţie pozitivă a clientului. De aceea, este important să acordăm o atenţie deosebită nivelului intensităţii VI asupra VD. De cele mai multe ori acest nivel trebuie ales, astfel încât să corespundă cât mai bine realităţii, pentru a avea o validitate ecologică sporită. Nu întotdeauna trebuie să mărim intensitatea VI pentru a observa efectul său asupra VD. Alteori avem nevoie de o analiză detaliată a evoluţiei VI pe toată gama amplitudinii sale. Acest fapt poate fi observat şi în legea Yerkes-Dodson. Dacă am dori să vedem care este efectul nivelului de arousal (motivare) asupra performanţei în sarcină şi am avea doar două grupe de subiecţi (slab motivaţi, respectiv puternic motivaţi) am putea ajunge la concluzia că nu există diferenţe între cele două grupe şi implicit, că nu există un efect al nivelului de motivare asupra performanţei subiecţilor. Dacă în loc de două grupe, am utiliza trei loturi, corespunzătoare a trei trepte ale VI (slab motivaţi, optim motivaţi, puternic motivaţi) am ajunge la rezultate diferite şi am admite influenţa nivelului de arousal asupra performanţei subiecţilor în sarcină.

6. *Tratarea variabilelor cauză drept factori ficşi*

Din fericire, de cele mai multe ori, VI sunt tratate automat ca factori ficşi în cadrul tehnicilor ANOVA. Deşi tratarea VI ca factor aleator poate facilita generalizarea rezultatelor obţinute la alte trepte ale VI, tehnica suferă de o putere statistică redusă. Într-un experiment tipic unifactorial, vom avea 2-3 trepte ale VI şi aproximativ 90 de subiecţi. Acest lucru este normal, dacă VI este tratată ca factor fix, însă dacă VI este considerată factor aleatoriu va fi nevoie, pe lângă un număr mare de subiecţi, de un număr ridicat de trepte ale VI pentru a creşte puterea statistică necesară detectării efectelor existente.

7. *Utilizarea unor ipoteze unidirecționale (unilaterale)*

În condițiile utilizării unei ipoteze de tip unidirecțional se dublează şansele de a obține rezultate semnificative statistic, deci creşte puterea statistică a cercetării. Astfel, dacă efectuăm o corelație pe 25 de subiecți între două probe şi obținem o valoare de -.30 vom interpreta diferit rezultatele în funcție de tipul de ipoteză avută. Datele ar fi sprijinit o ipoteză unidirecțională („persoanele cu un nivel crescut de depresie tind să petreacă mai puțin timp în oraş?"), r = -.30, p < .05, însă ar fi infirmat o ipoteză bidirecțională („există vreo asociere între nivelul de depresie a unei persoane şi perioada de timp petrecută în oraş?"), r = -.30, p > .05.

8. *Utilizarea scalelor de măsurare numerice, de tip interval sau proporții*

Acest gen de scale de măsurare apelează frecvent la tehnici statistice parametrice. Din perspectiva puterii statistice este cunoscut faptul că tehnicile statistice parametrice au un nivel mai ridicat decât metodele statistice neparametrice. În consecință, ar fi de dorit ca cel puțin variabilele dependente sau criteriu să fie măsurate prin scale numerice, evitându-se, pe cât posibil, măsurarea lor prin scale de tip nominal sau ordinal.

9. *Transformarea datelor pentru a îndeplini condițiile diferitelor tehnici statistice*

Chiar dacă datele sunt măsurate prin scale numerice, majoritatea tehnicilor statistice necesită alte condiții suplimentare pentru a putea fi aplicate. Nerespectarea asumpțiilor majore, în situații de aplicare mecanică a testelor statistice, poate conduce la erori de tip I sau II în testarea ipotezelor. Spre exemplu, o distribuție complet anormală a variabilei dependente ar presupune renunțarea la aplicarea testului t pentru două eşantioane independente, în favoarea unei metode neparametrice corespunzătoare, în acest caz, testul U al lui Mann-Whitney. Puterea statistică a metodei este însă la nivelul a 69% din capacitatea testului t de a detecta diferențe între cele două grupe. În asemenea condiții pot fi adoptate o serie de artificii statistice de transformare a datelor brute, menite să corecteze distribuția obținută, astfel încât să nu fie probleme cu aplicarea testelor parametrice. Există o multitudine de transformări posibile, cele mai întâlnite fiind ridicarea la pătrat, extragerea rădăcinii pătrate sau utilizarea logaritmilor naturali. Iată, spre exemplu, valorile indicatorilor de asimetrie şi boltire, înainte şi după transformare, dintr-o bază de date fictivă (tabelul 2.4):

Tabelul 2.4. Valori descriptive înainte de transformare X şi după transformare X^2

Variabilă	Asimetrie	Eroare standard	Boltire	Eroare standard
X	-.978	.314	.082	.618
X^2	-.439	.314	-.952	.618

Înainte de transformarea datelor se observă o deviere importantă de la normalitate, fiind o distribuție puternic asimetrică negativ. De altfel, anormalitatea distribuției este susținută şi de rezultatul testului Kolmogorov-Smirnov (z = 1,50, p < .05). După ridicarea la pătrat a rezultatelor se reduce foarte mult asimetria datelor, însă intervin unele modificări la nivelul boltirii. Totuşi, după transformare, distribuția obținută nu diferă semnificativ de una normală (z = 1,13, p > .05), fapt ce va permite aplicarea unor teste parametrice. Prin asemenea transformări scad şansele de a comite erori de tip I sau II. O parte din transformările posibile au fost discutate în secțiunea 1.4.3. din capitolul vizând statistica descriptivă.

10. *Selectarea unor probe cu caracteristici psihometrice ridicate, mai ales în cazul măsurării VD*

Itemii care au un nivel de eroare scăzut şi care reuşesc să reflecte cu sensibilitate orice variații ale variabilei latente corespunzătoare conduc la o putere statistică mai mare decât cei care nu îndeplinesc aceste condiții. Spre exemplu, o scală care are o consistență internă de .65, are un nivel ridicat al erorii, fapt ce afectează relaționarea sa cu alte probe. De pildă, în corelarea scalei cu alte probe se vor obține valori scăzute, sub nivelul a .65, deoarece ea nu poate obține corelații mai mari decât cu ea însăşi (r = .65). Astfel, relația sa cu o altă variabilă este atenuată şi poate deveni chiar nesemnificativă statistic. Spre exemplu, am putea obține o corelație de .25, care ar fi fost mai mare dacă ambele probe ar fi avut o consistență internă ridicată, în jurul a .90. În privința sensibilității, teoria răspunsului la itemi oferă un bun mijloc de a testa şi selecta acei itemi care răspund cel mai bine cerinței de sensibilitate. Nu în ultimul rând, scalele care corelează moderat sau puternic cu probele de dezirabilitate socială pot duce la scăderea puterii statistice.

11. *Salvarea unor grade de libertate*

Deşi am amintit implicit de acest aspect la punctul patru, salvarea unor grade de libertate, cu scopul de a spori puterea unei cercetări, poate fi făcută şi într-un design complex (de exemplu, ANOVA 3x3x3). Într-un asemenea caz am putea calcula şapte indicatori F, câte trei pentru efectele simple (F_A, F_B, F_C), pentru interacțiunile de ordinul doi (F_{AB}, F_{AC}, F_{BC}) şi unul pentru interacțiunea complexă de ordinul trei (F_{ABC}). Nu are rost să calculăm toate aceste efecte în condițiile în care sunt interacțiuni care nu ne interesează sau în cazul în care interpretarea rezultatelor ar fi dificilă sau lipsită de sens (ex. F_{ABC}). Renunțând la calcularea acelor indici, vom salva un anumit număr de grade de libertate, ceea ce ar duce la o creştere uşoară a puterii statistice a testelor F rămase a fi efectuate. În situații limită, prin salvarea unor grade de libertate, rezultatele nesemnificative pot deveni semnificative (de exemplu, un p de .07 se schimbă într-un p de .04).

12. *Utilizarea tehnicilor statistice cu putere ridicată*

În practică, putem întâlni situații în care pot fi aplicate mai multe tehnici statistice. De exemplu, în cadrul comparațiilor post hoc din ANOVA, pentru grupe egale de subiecți şi omogenitatea dispersiei asumată, am putea alege între procedurile Tukey şi Bonferroni. Totuşi, primele au o putere mai mare de a detecta diferențe între grupe în comparație cu ultimele. Similar, într-un design intragrup putem apela la ANOVA pentru măsurători repetate în variantă corectată sau putem alege MANOVA. Prima opțiune are o putere statistică mai mare decât MANOVA, de aceea ar fi preferabilă utilizarea sa.

Toate aceste măsuri pot fi utilizate în sprijinul creşterii puterii statistice a unei cercetări. Vom încheia precizând că problema prezentată nu este doar un moft al statisticienilor, menit să îndepărteze şi mai mult comunitatea psihologică de la o abordare ştiințifică a problemelor. APA (2001) oferă dovada cea mai clară în acest sens şi face recomandarea de a se ține cont de puterea statistică în stabilirea designului şi în interpretarea rezultatelor unei cercetări.

2.4. Appendix – Ce înseamnă gradele de libertate?

Cei care s-au întrebat vreodată ce înseamnă gradele de libertate şi au primit răspunsuri evazive ori prea elevate merită să citească secțiunea de mai jos. Gradele de libertate însoțesc întotdeauna rezultatele statisticii inferențiale. La recomandările APA (2001), în articolele de specialitate se regăsesc exprimări de genul $t(23) = 4{,}11$, $p < .01$, ori $F(2, 20) = 1{,}73$, $p > .05$. Dacă t sau F reprezintă simboluri ale tehnicilor statistice utilizate, iar p este pragul de semnificație asociat acelor valori, datele din paranteză înseamnă coordonatele gradelor de libertate. În varianta clasică, aceste coordonate ne ajutau să descoperim valorile statistice din tabel care trebuie comparate cu valorile calculate de noi. O dată cu apariția statisticii pe calculator, acest lucru a devenit un atavism, deoarece pragul de semnificație este oferit de calculator, împreună cu rezultatele obținute.

Totuşi, ce sunt gradele de libertate şi care sunt elementele care determină valoarea lor? Cu umor, Yu (2003) remarca faptul că aceste grade de libertate nu au nimic de a face cu căsătoria, deşi un bărbat însurat, având o singură parteneră, nu mai are nici un grad de libertate. Acelaşi autor oferă o excelentă explicație intuitivă pentru gradele de libertate pornind de la elementele esențiale ce intervin în determinarea lor: volumul eşantionului, respectiv numărul de parametri care trebuie calculați. Fără a avea caracter de lege, gradele de libertate pot fi exprimate frecvent prin numărul de observații minus numărul de relații necesare pentru a obține acele date (n-r).

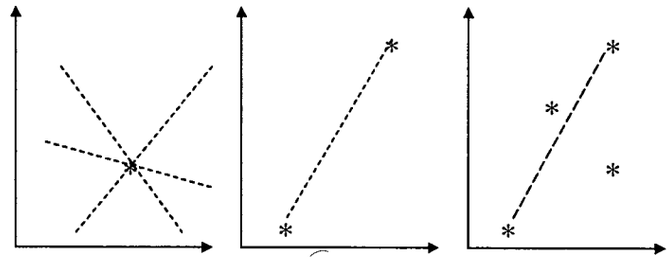


Figura 2.3. *Dependența gradelor de numărul de observații (zero grade de libertate – stânga; potrivirea perfectă – în mijloc; suprapotrivirea – în dreapta)*

În imaginea din stânga avem un singur caz, adică un singur subiect testat. Dacă dorim să estimăm direcția sa de evoluție, vom avea o multitudine de posibilități (liniile punctate). Pornind de la ideea că gradele de libertate pot fi reduse la expresia n-r, în cazul de față vom obține valoarea 0, deoarece avem un singur subiect (n=1) şi o singură estimare de făcut, vizând evoluția ulterioară a subiectului (r = 1). Într-o situație cu zero grade de libertate nu există nici o posibilitate de a dovedi falsitatea sau veridicitatea unui model, deoarece orice soluție poate fi sprijinită.

În imaginea din mijloc, avem două observații şi o singură relație de demonstrat (gradul de asociere dintre cele două observații). Făcând calculul, ne aflăm într-o situație cu un singur grad de libertate. De această dată vom avea o singură soluție corectă, estimată prin linia punctată care trece prin cele două puncte. Cazul se numeşte „potrivire perfectă" (în engleză „perfect fit"), însă în ciuda numelui, nu ne dorim o asemenea situație. De ce? Deoarece modelul găsit perfect este singurul posibil. Orice alt model alternativ nu poate fi comparat, pentru a studia în ce măsură modelul inițial descrie cel mai bine şirul de date.

În imaginea din dreapta situația se schimbă uşor. Punctele din imagine nu mai reprezintă observații, ci variabile. Dacă numărul variabilelor de estimat este mai mare decât numărul de observații, ar trebui să obținem o valoarea negativă a gradelor de libertate. Această situație nedorită se numeşte suprapotrivire (în engleză „over-fitting").

Recapitulând, la zero grade de libertate nu putem determina o soluție; în caz de potrivire perfectă, soluția este întotdeauna adevărată; iar în situații de suprapotrivire, soluția este aproape întotdeauna adevărată. De aceea întâlnim puține cercetări bazate pe tehnici statistice inferențiale, efectuate pe un număr mic de subiecți, iar atunci când se întâmplă, rezultatele lor sunt privite cu circumspecție. Pentru a elimina neajunsurile prezentate, va trebui să avem mult mai multe observații, decât parametrii de estimat. Numai astfel vom putea evalua cu încredere calitatea (falsitatea) unui modelul testat.

BIBLIOGRAFIE

APA (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

Becker, L. (2003). *Effect size*. Găsit la: http://www.uccs.edu/~lbecker/psy590/es.htm.

Clocotici, V., Stan, A. (2000). *Statistică aplicată în psihologie*. Iaşi: Editura Polirom.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.

Cohen, J. (1994). The earth is round (p< .05). *American Psychologist*, 49(12), 997-1003.

Dyer, C. (1995). *Beginning research in psychology: a practical guide to research methods and statistics*. Oxford, UK: Blackwell Publishers Ltd.

Embertson, S.E., Reise, S.P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates, Publishers.

Fan, X. (2001). Statistical significance and effect size in educational research: two sides of a coin. *The Journal of Educational Research*, 94(5), 275-282.

Faul, F., Erdfelder, E. (1992). *GPOWER: A priori, post hoc, and compromise power analyses for MS-DOS [Computer program]*. Bonn, Germany: Bonn University, Dept. of Psychology.

Hastie, R., Stasser, G. (2000). Computer simulation methods for social psychology. În H.T. Reis & C.M. Judd (eds.). *Handbook of research methods in social and personality psychology* (p.85-114). Cambridge: Cambridge University Press.

Hinkle, D.E., Wiersma, W., Jurs, S.G. (1994). *Applied statistics for the behavioral sciences* (3rd ed.). Boston: Houghton Mifflin Company.

Kirk, R.E. (1982). *Experimental design: procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks / Cole Publishing.

Leventhal, L., Huynh, C.L. (1996). Directional decisions for two-tailed tests: power, error rates, and sample size. *Psychological Methods*, 1(3), 278-292.

Pedhazur, E.J., Schmelkin, L.P. (1991). *Measurement, design, and analysis: an integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Rosenthal, R., DiMatteo, M.R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.

Rosenthal, R., Rosnow, R.L., Rubin, D.B. (2000). *Contrasts and effects sizes in behavioral research. A correlational approach*. Cambridge: Cambridge University Press.

Sava, F. (2001). Reducing teachers' misbehaviour: an action approach. *Revista de Psihologie Aplicată*, 3(4), 31-42.

Sava, F. (2002). Ten ways to increase statistical power. *Analele Universităţii de Vest din Timişoara. Seria Sociologie, Psihologie, Pedagogie şi Asistenţă Socială*, 8, 89-92.

Sava, F. (2003). Puterea statistică şi testarea ipotezelor. *Psihologia Resurselor Umane*, 1(2), 28-34.

Singer, B.R., Lovie, A.D., Lovie, P. (1986). Simple size and power. În A.D. Lovie (ed.). *New developments in statistics for psychology and social sciences* (p.129-142). London: The British Psychological Society and Methuen.

Smith, E.R. (2000). Research design. În H.T. Reis & C.M. Judd (eds.). *Handbook of research methods in social and personality psychology* (p.17-39). Cambridge: Cambridge University.

Thompson, B. (1999). If statistical significance tests are broken / misused, what practices should supplement or replace them? *Theory & Psychology*, 9(2), 165-181.

Thompson, B. (2002). „Statistical", „practical", and „clinical": how many kinds of significance do counselors need to consider? *Journal of Counseling and Development*, 80, 64-71.

Tversky, A., Kahneman, D. (1993). Belief in the law of small numbers. In G. Keren & C. Lewis (eds.). *A handbook for data analysis in the behavioral sciences: methodological issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

West, S.G., Biesanz, J.C., Pitts, S.C. (2000). Causal inference and generalization in field settings: experimental and quasi-experimental designs. In H.T. Reis & C.M. Judd (eds.). *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press, 40-84.

Yu, C.H.A. (2003). *Illustrating degrees of freedom in terms of simple size and dimensionality*. Găsit la adresa http://seamonkey.ed.asu.edu/~alex/computer/sas/df.html.

CAPITOLUL 3

TEHNICI INTRODUCTIVE DE TESTARE A IPOTEZELOR

Dacă în capitolul anterior am insistat asupra logicii şi principiilor de bază în testarea ipotezelor, în secțiunea următoare vom aplica cunoştințele respective la nivelul celor mai răspândite tehnici statistice introductive. Testarea ipotezelor, în cazul unui studiu cu două variabile, poate fi înțeleasă simplificat ca răspunsuri la una dintre următoarele întrebări: „Există o asociere semnificativă statistic între două variabile?" sau „Există diferențe semnificative statistic între două grupe?". De notat că sintagma „semnificativă statistic" nu se referă la amplitudinea corelației sau a diferențelor observate, ci implică ideea unei probabilități foarte mici de a obține datele culese în condițiile în care ipoteza nulă este adevărată. Premisele acesteia indică, fie că „nu există asociere între cele două variabile", fie că „nu există diferențe între cele două populații din care au fost extrase eşantioanele".

3.1. Testarea asocierii dintre două variabile

3.1.1. Coeficientul de corelație r

Prin analiza corelațională se urmăreşte în ce măsură schimbările dintr-o variabilă sunt însoțite de schimbări într-o altă variabilă. Coeficientul de corelație liniară r este cel mai întâlnit indicator de măsurare a gradului de asociere dintre două variabile. Aşa cum am precizat în capitolul destinat prezentării elementelor de statistică descriptivă, r poate varia între -1 şi 1, unde 0 semnifică absența legăturii dintre două variabile.

Să presupunem că între nivelul de extroversiune şi nivelul de neuroticism nu există legătură, adică avem un r egal cu 0. Pentru a testa relația, vom extrage un eşantion de 30 de persoane din cadrul studenților la psihologie, vom testa nivelul celor două dimensiuni ale personalității şi vom obține o valoare r de -.09. Dacă am fi testat aceeaşi relație pe alți 30 de studenți la psihologie, aleşi la întâmplare, am fi obținut un r de .04. Extrăgând alte loturi de subiecți, vom observa că majoritatea indicatorilor r sunt apropiați de zero şi mult mai puțini sunt îndepărtați de această valoare. Din păcate, în practică nu ne permitem să extragem atât de multe eşantioane pentru a testa relația dintre două variabile. De obicei se testează un singur lot de persoane, rezultând o valoare r unică, să zicem -.12. Întrebarea devine: ce probabilitate există să obținem un r de -.12 în condițiile în care nu există legătură între cele două variabile (r real este 0)? Indică un r de

-.12 absența asocierii între cele două variabile? Conform statisticii descriptive nu ne putem da seama de acest lucru, deoarece valoarea de -.12 poate fi interpretată ca o uşoară asociere negativă între cele două variabile, ceea ce implică faptul că persoanele cu scoruri ridicate la extroversiune tind să aibă un nivel mai scăzut de neuroticism.

Din păcate, valorile descriptive sunt dependente de o serie de factori, în special de volumul eşantionului. Un r de -.12 indică o legătură negativă între cele două variabile în cazul unui eşantion de 1000 de persoane, respectiv absența relației dintre variabile în cazul unui eşantion de 50 de persoane. Pentru a lămuri această dilemă este nevoie să privim valoarea coeficientului de corelație din perspectiva statisticii inferențiale.

Iată un exemplu ipotetic, menit să faciliteze interpretarea inferențială a lui r:

Conform unei analize efectuate de Leary, Tchividjian şi Kraxberger (1994/1999), între nivelul de auto-prezentare (management al impresiei) şi cheltuielile lunare pe produse cosmetice există o asociere pozitivă. Un nivel ridicat de management al impresiei indică un autocontrol puternic față de modul în care o persoană este percepută de către cei din jur şi implicit, mai multă grijă față de aspectul său fizic. De aceea, ne aşteptăm ca un nivel crescut de auto-prezentare să fie asociat cu mai multe cheltuieli pe produse cosmetice. Pentru a studia relația dintre cele două variabile, vom colecta date cu privire la cheltuielile lunare, exprimate în EURO, pe produse cosmetice şi la nivelul de management al impresiei, obținut pe baza unui chestionar. Lotul testat include 50 de adolescenți (25 de fete şi 25 de băieți), ale căror răspunsuri sunt prezentate în baza de date „*corelatii.sav*".

Pentru a rezolva problema, va trebui să răspundem la următoarele întrebări:

1) Cum identificăm faptul că este vorba de o corelație?
2) Ce fel de date au fost colectate?
3) Condițiile aplicării corelației respective sunt îndeplinite?
4) Ce fel de ipoteză trebuie testată?
5) Cât de intensă este asocierea dintre cele două variabile?

Răspunsurile la primele patru întrebări ne ajută să găsim coeficientul de corelație care trebuie calculat şi ne oferă o imagine despre gradul de adecvare a interpretării valorii găsite. Ultima întrebare vizează un aspect distinct, ce ține de mărimea efectului, adică de intensitatea relației dintre cele două variabile.

(1) Cum identificăm faptul că este vorba de o corelație?

Multe dintre studiile întreprinse în psihologie apelează la metoda corelației. Putem identifica necesitatea analizei corelaționale din modul de formulare a obiectivului, respectiv a ipotezei de cercetare. Expresii de genul: „urmărim să evidențiem *gradul de asociere* dintre inteligență şi memorie", „*există o legătură*

între agresivitate şi frustrare", „temperaturile crescute *tind* să conducă la un consum ridicat de lichide", „vom analiza *relația* dintre extroversiune şi timiditate" indică ideea de corelație. În exemplul dat, expresiile „un nivel crescut de auto-prezentare este asociat cu un consum mai mare de produse cosmetice" ori „pentru a studia relația dintre cele două variabile" sunt indiciile care ne conduc la ideea necesității de a calcula un coeficient de corelație.

(2) Ce fel de date au fost colectate?

Cheltuielile lunare pe produse cosmetice, exprimate în euro, reprezintă date numerice, de tip proporții, în timp ce nivelul de management al impresiei este măsurat prin scale numerice de tip interval, prin cumularea răspunsurilor la cele cinci întrebări. Astfel, cheltuielile pe produse cosmetice sunt cuprinse între 2 şi 28 de EURO, iar nivelul de management al impresiei este situat între 5 şi 20 de puncte, un rezultat ridicat însemnând un nivel crescut de auto-prezentare. Ținând cont că ambele variabile sunt măsurate prin scale numerice, vom apela, cel mai probabil, la coeficientul de corelație r.

(3) Condițiile aplicării lui r sunt îndeplinite?

Pe lângă caracterul numeric al datelor colectate, pentru a aplica r mai este nevoie ca relația dintre cele două variabile să fie liniară şi ca norul de puncte să indice o situație de homoscedasticitate. Ambele aspecte, discutate în capitolul introductiv de statistică descriptivă, pot fi apreciate prin inspectarea norului de puncte din figura 3.1.

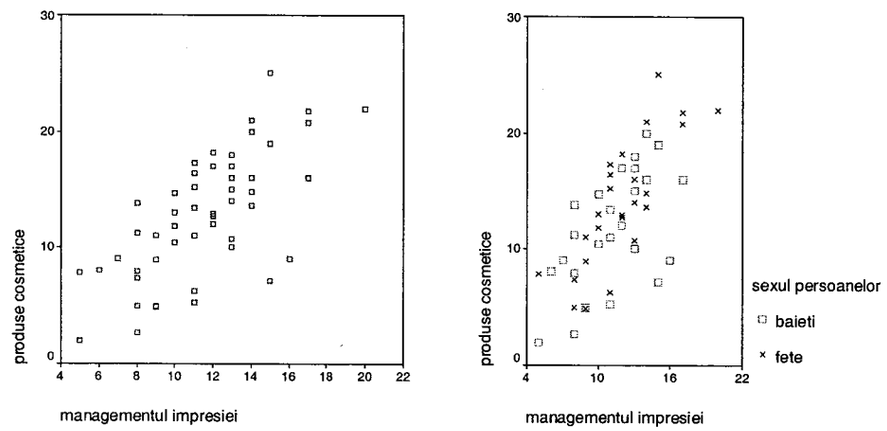


Figura 3.1. *Norul de puncte dintre cele două variabile nediferențiat (în stânga), respectiv diferențiat în funcție de sexul respondentului (în dreapta)*

Dacă analizăm imaginea din partea stângă, vom constata atât liniaritatea relației (punctele par a evolua dinspre stânga jos spre dreapta sus, indicând o

asociere liniară pozitivă între variabile), cât şi homoscedasticitatea relației (punctele par a forma un nor de puncte de grosime similară pe întreaga amplitudine a evoluției celor două variabile). Prin urmare, toate condițiile aplicării lui r sunt îndeplinite. Totuşi, dacă vom analiza şi imaginea din dreapta, cu norul de puncte rezultat în urma diferențierii dintre cele două sexe ale subiecților respondenți, vom observa că, în cazul băieților, punctele sunt ceva mai împrăştiate (norul de puncte fiind mai gros). Această imagine ne face să credem că intensitatea legăturii dintre cele două variabile este uşor diferită pentru fete, respectiv băieți. Cu alte cuvinte, ne aşteptăm să existe o asociere mai mare între preocuparea pentru managementul impresiei şi cheltuielile lunare pentru produse cosmetice în cazul fetelor, decât în cazul băieților. De aceea, pe lângă valoarea lui r pentru întregul lot de subiecți, vom calcula şi valorile r pentru fiecare sex în parte. Această măsură este complementară în cazul corelației, însă devine esențială în stabilirea unor ecuații de regresie predictive.

Tot din imaginea respectivă se observă cheltuieli mai ridicate din partea fetelor pentru produse cosmetice în comparație cu cele ale băieților.

(4) Ce fel de ipoteză trebuie testată?

Ipotezele de cercetare pot fi exprimate, fie în termeni unidirecționali (unilaterali), fie în termeni bidirecționali (bilaterali). În prima categorie includem expresii de genul „persoanele cu un nivel de inteligență ridicat tind să obțină performanțe superioare în rezolvarea problemelor de logică" sau „un nivel crescut de anxietate tinde să fie asociat cu un nivel scăzut de agresivitate verbală". În ambele expresii este precizată clar direcția asocierii dintre variabile, ea fiind pozitivă în primul caz („când X creşte, Y tinde să crească") şi negativă în cel de-al doilea („când X creşte, Y tinde să scadă"). În schimb, ipotezele bidirecționale nu au precizată clar direcția de asociere dintre două variabile. Expresii de genul „există o corelație între X şi Y" sau „există o corelație între problemele de sănătate şi nivelul de management al impresiei" sunt tipice pentru cazul ipotezelor bidirecționale.

În exemplul nostru putem distinge situația unei ipoteze unilaterale, deoarece direcția asocierii dintre variabile este precizată. Cercetătorul se aşteaptă ca un nivel crescut de autoprezentare să fie asociat cu un consum mai mare de produse cosmetice.

Importanța distingerii dintre cele două tipuri de ipoteze se regăseşte în valorile r semnificative statistic. Testarea ipotezelor unidirecționale urmăreşte în ce măsură valoarea lui r este semnificativ diferită de 0 în sens pozitiv (în cazul unei prezise asocieri pozitive), respectiv în sens negativ (în cazul unei prezise asocieri inverse). În schimb, testarea ipotezelor bidirecționale urmăreşte în ce măsură r diferă semnificativ de 0, în oricare direcție, fie spre -1, fie spre +1. Datorită acestor diferențe, există de două ori mai multe şanse să obținem un rezultat semnificativ statistic în cazul ipotezelor unidirecționale decât în cazul celor bidirecționale.

Sintetizând răspunsurile cheie pentru problema dată, vom concluziona că putem aplica coeficientul de corelație r (coeficientul de corelație r Bravais-

Pearson), deoarece ipoteza doreşte stabilirea gradului de asociere dintre preocuparea pentru managementul impresiei şi cheltuielile lunare pentru produse cosmetice. Ambele variabile sunt măsurate prin scale numerice, iar relația dintre ele este liniară şi homoscedastică. În plus, pentru testarea valorii r va trebui să ținem seama de faptul că ipoteza este formulată în termeni unidirecționali (unilaterali).

În SPSS vom obține valoarea lui r, dacă vom urma calea din meniu: „Analyze"- „Correlate"-„Bivariate". Fereastra deschisă ca urmare a acestor opțiuni are selectată ca variantă standard „Pearson", adică exact coeficientul de corelație de care avem nevoie. Ne rămâne doar să selectăm variabilele ce urmează a fi corelate şi să modificăm opțiunea „two-tailed" (bidirecțional) în „one-tailed" (unidirecțional). Output-ul oferit de SPSS este prezentat în figura 3.2.

		produse cosmetice	managementul impresiei
produse cosmetice	Pearson Correlation	1	,699**
	Sig. (1-tailed)	,	,000
	N	50	50
managementul impresiei	Pearson Correlation	,699**	1
	Sig. (1-tailed)	,000	,
	N	50	50

**. Correlation is significant at the 0.01 level (1-tailed).

Figura 3.2. *Output-ul SPSS cu rezultatul corelației dintre cele două variabile*
SPSS este o marcă înregistrată.

Rezultatul se poate scrie simplu sub forma: $r(48) = .70$, $p < .01$. În expresie, r este simbolul coeficientului de corelație liniară simplă, iar între paranteze sunt trecute gradele de libertate ale testului. Valoarea 48 indică gradele de libertate, care se calculează, în cazul corelației, prin formula N-2, unde N semnifică numărul de subiecți testați. Valoarea de .70 este rezultatul rotunjit observat la intersecția dintre managementul impresiei şi produsele cosmetice, adică r. În sfârşit, $p < .01$ semnifică probabilitatea de a obține datele observate în condițiile în care nu există o asociere între managementul impresiei şi nivelul cheltuielilor pentru produse cosmetice. În cazul de față se observă o probabilitate mai mică de 1 la 100 (< .01) de a obține un r de .70, în condițiile în care nu ar exista o corelație între cele două variabile.

Prin urmare, datele sprijină ipoteza lansată, conform căreia o preocupare crescută pentru crearea unei impresii favorabile în ochii celorlalți se asociază cu cheltuieli mai ridicate pentru achiziționarea de produse cosmetice. Rămâne să stabilim magnitudinea relației dintre cele două variabile, nu înainte de a aminti că relația dintre cele două variabile este $r(23) = .79$, $p < .01$ în cazul fetelor, respectiv $r(23) = .58$, $p < .01$ în cazul băieților.

(5) Cât de intensă este asocierea dintre cele două variabile?

Majoritatea textelor statistice evită să răspundă la această întrebare, rezumându-se să prezinte rezultatele sub forma menționată anterior. Totuşi, aşa cum am subliniat în capitolul destinat testării ipotezelor, pragul de semnificație p nu indică intensitatea unei relații, ci raritatea de a obține asemenea date în condițiile în care ipoteza nulă este adevărată. Spre exemplu, la un număr de 700 de persoane testate putem obține un r foarte mic, dar semnificativ statistic, cum ar fi r (698) = .08, p < .001.

Cel mai utilizat criteriu pentru a aprecia intensitatea relației dintre două variabile se numeşte coeficient de determinare, notat cu r^2. Aşa cum arată simbolul său, se calculează foarte uşor, fiind pătratul coeficientului de corelație. El poate fi exprimat şi în procente, unde 100% corespunde unui coeficient de corelație de -1 sau 1. Din păcate, acest criteriu nu este întotdeauna adecvat din cauza influenței pe care o are mărimea lotului în determinarea valorii lui r şi, implicit, a coeficientului de determinare. El trebuie analizat cu grijă în cercetările corelaționale cu mai puțin de 20 de subiecți, respectiv cele care depăşesc 200 de persoane.

Prin intermediul coeficientului de determinare se evaluează partea de asociere comună a celor două variabile, indicând procentul din dispersia unei variabile care se asociază cu modul de împrăştiere a celeilalte variabile. Cu alte cuvinte, coeficientul de determinare arată gradul de comunalitate dintre două variabile. În cazul în care una dintre variabile este independentă, iar cealaltă dependentă, interpretarea se modifică uşor. Coeficientul de determinare poate fi înțeles ca procentul din dispersia variabilei dependente explicat de evoluția (influența) variabilei independente (Pedhazur şi Schmelkin, 1991). Cu alte cuvinte, coeficientul de determinare ne arată cât de mare este influența variabilei independente asupra variabilei dependente.

În exemplul ilustrat, vom obține un coeficient de determinare r^2 de .49 (49%), ceea ce înseamnă că aproape jumătate din dispersia celor două variabile are o evoluție comună. De obicei, coeficientul de determinare atinge valori mult mai mici. Prin comparație cu criteriul orientativ stabilit de Cohen (1988) pentru indicele d, putem aprecia că un coeficient de determinare de .01 indică o asociere scăzută între două variabile, o valoare de .06 indică o asociere de intensitate medie între două variabile, iar un r^2 de .14 semnifică o asociere puternică între două variabile. În consecință, între gradul de preocupare pentru managementul impresiei şi cheltuielile lunare pentru produse cosmetice există o legătură foarte puternică. Cu cât gradul de preocupare pentru imaginea avută în ochii celorlalți este mai mare, cu atât există o tendință de a cheltui mai mult pe produse cosmetice.

O altă limită a coeficientului de determinare o constituie valoarea practică a unei asocieri. Uneori, aşa cum am amintit de relația dintre aspirină şi infarctul miocardic, un coeficient de determinare extrem de mic, de numai 0,1%, arată că 34 din 1000 de atacuri de cord ar fi putut fi evitate prin consumul de aspirină. Deci, pe lângă magnitudinea schimbării evidențiată prin indicatori ai mărimii efectului, precum r^2, va trebui să ținem seama şi de aspectul valoric al schimbării.

În fine, se cuvine să reamintim că există o convenție de folosire a semnelor în cazul indicatorilor de mărime a efectului. Aceştia vor lua valori negative în cazul în care efectul apare în direcția opusă celei precizate în ipoteză, respectiv semnul plus, dacă efectul apare în direcția scontată.

Limitele interpretării coeficientului de corelație

Până la urmă, corelația nu este decât un procedeu statistic ce măsoară covarianța dintre două variabile standardizate. Totuşi, din punct de vedere metodologic, calitatea unui coeficient de corelație poate diferi, deşi indică aceeaşi cantitate a asocierii. Spre exemplu, un r de .60 arată un grad de covarianță ridicat între două variabile, deşi sursa legăturii dintre ele poate fi diferită. Iată câteva idei de bază:

Corelația nu implică o cauzalitate

Mulți începători tind să testeze, prin metoda corelației, ipoteze ce conțin sintagme cauzale. O expresie de genul „în ce măsură inteligența influențează performanțele şcolare ale subiecților" nu poate fi demonstrată prin intermediul lui r. Valoarea coeficientului de corelație poate arăta doar dacă cele două variabile covariază în aceeaşi direcție, însă un r semnificativ nu poate fi un substitut pentru a demonstra existența unei relații cauzale. Aceasta poate fi demonstrată, fie printr-un design experimental ce presupune manipularea variabilelor de interes de către cercetător, fie printr-un demers logic, în care relația dintre cele două variabile este clară (căldura excesivă duce la creşterea consumului de lichide şi nu creşterea consumului de lichide duce la creşterea temperaturii aerului).

Sursa covarianței dintre două variabile poate fi diversă

Brewer (2000) oferă o sinteză excelentă cu privire la posibilele surse ale corelației dintre două variabile. Acestea au un rol deosebit în stabilirea validității interne a unei cercetări.

O posibilă cauză a relației dintre două variabile este tocmai dependența cauzală a uneia față de cealaltă, cum ar fi creşterea consumului de lichide cu cât gradul de deshidratare creşte.

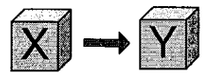


Figura 3.3. *Schema unei relații cauzale între două variabile*

O altă cauză posibilă a relației dintre două variabile este asocierea (dependența) uneia sau ambelor variabile de o terță variabilă. Asemenea asociații se numesc corelații false (în engleză „spurious correlations").

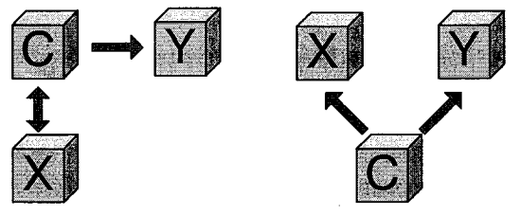


Figura 3.4. *Schema unei relații false între două variabile X şi Y*

În partea stângă a figurii 3.4, corelația dintre X şi Y este rezultatul influenței unei variabile terțe C. Aceasta are o relație non-cauzală cu X şi una cauzală cu Y. Să presupunem că un cercetător ajunge la concluzia că zilele ploioase (X) duc la o dispoziție afectivă depresivă (Y). O asemenea concluzie este eronată, deoarece o mulțime de fenomene atmosferice, precum presiunea aerului sau gradul de luminozitate, sunt asociate cu prezența sau absența ploii fără a exista o relație de cauzalitate între ele. În schimb, aceste condiții atmosferice pot influența dispoziția afectivă a subiecților. În consecință, dispoziția afectivă nu este dependentă de prezența sau absența ploii, ci de condițiile atmosferice care însoțesc acest fenomen natural. De remarcat că pot fi o multitudine de variabile terțe care să afecteze relația dintre două variabile.

În partea dreaptă a figurii 3.4, cele două variabile X şi Y ajung să varieze deoarece ambele sunt influențate de o variabilă terță. O situație tipică ar fi relația testată pe un lot de copii cuprinşi între 2 şi 7 ani, dintre numărul de ore de somn şi numărul de cuvinte ştiute. Corelația falsă dintre cele două variabile ne-ar putea duce la concluzia că o persoană care doarme mai puțin, ştie mai multe cuvinte. Situația ar putea fi uşor remediată dacă am conştientiza şi controla rolul vârstei asupra celor două variabile, ştiindu-se dependența evoluției somnului şi a dezvoltării limbajului de vârsta cronologică.

Variabilele ascunse (neluate în seamă) nu sunt singurele care pot influența validitatea internă a rezultatelor unei cercetări. Chiar şi în situația existenței unei relații cauzale, aceasta poate fi moderată de prezența altor variabile. Moderarea relației presupune creşterea sau descreşterea intensității asocierii dintre două variabile. Un exemplu cotidian de variabilă moderatoare este tipul de ten, care afectează relația dintre perioada de stat la soare şi nivelul de bronz al pielii. În psihologie, multe variabile de tip etichetă, precum sexul persoanei, vârsta, nivelul de experiență pot influența relația dintre două variabile. Cel mai recent exemplu poate fi observat în imaginea din dreapta a figurii 3.1, în care se observă că sexul unei persoane moderează relația dintre managementul impresiei şi volumul cheltuielilor pentru produse cosmetice. Astfel, în cazul fetelor relația dintre cele

două variabile este mai intensă (r = .79) decât în cazul băieților (r = .58). Uneori, variabilele moderatoare pot chiar bloca relația cauzală dintre X şi Y. Astfel, am putea concluziona eronat că timpul petrecut la soare nu influențează nivelul de bronzare a unei persoane. Datele culese indică acest lucru, deoarece toți subiecții testați au fost îmbrăcați în echipamente de protecție. Acestea joacă rolul unei variabile moderatoare a relației dintre X şi Y.

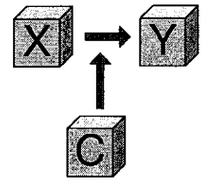


Figura 3.5. *Schema unei relații dintre X şi Y moderată de o terță variabilă*

În sfârşit, merită făcută distincția dintre variabilele moderatoare şi cele mediatoare. În cazul ultimelor, prezența variabilei terțe este necesară pentru apariția relației dintre X şi Y. Astfel, X afectează variabila C, care la rândul ei afectează variabila Y. În cazul relației dintre prezența sau absența ploii şi dispoziția afectivă, poate fi oferită o explicație alternativă, pe baza unei variabile mediatoare. Atunci când plouă (X), oamenii preferă să rămână în casă (C). Or, aceasta implică un număr redus de contacte sociale (C) şi poate favoriza apariția unei dispoziții afective negative (Y). Un alt exemplu, din psihologia educațională, priveşte influența concepției dascălului cu privire la calitatea relației profesor-elev. Aceasta la rândul ei, afectează motivația şcolară a elevilor.

Figura 3.6. *Schema unei relații dintre X şi Y mediată de o variabilă terță*

Toate aceste situații arată limitele validității asocierii dintre două variabile care nu pot fi izolate de influența altor aspecte neluate în studiu. De aceea, un concept interesant şi important este acela de corelație parțială.

3.1.2. Corelația parțială

Coeficientul de corelație dintre două variabile poate induce în eroare sau poate fi dificil de interpretat atunci când relația dintre cele două este afectată de o a treia. Acest lucru este evident mai ales în cazurile unor relații false sau a celor moderate. Pentru a elimina aceste neajunsuri se apelează la un artificiu statistic denumit corelație parțială.

Luând exemplul corelației dintre nivelul de management al impresiei şi cheltuielile lunare destinate produselor cosmetice, vom aprecia probabil, că relația este moderată, printre altele, de nivelul câştigurilor lunare ale părinților. Utilizând baza de date „corelatii partiale.sav" şi urmând calea „Analyze"-„Correlate"-„Partial" vom deschide fereastra din SPSS specifică indicatorului de corelație parțială. Înainte de a efectua această operație putem calcula corelațiile simple r dintre fiecare pereche de variabile. Astfel, vom obține r (48) = .70, p < .01 pentru relația dintre managementul impresiei şi cheltuielile lunare destinate produselor cosmetice, un r (48) = .51, p < .01, între managementul impresiei şi venitul familial lunar al subiecților, respectiv un r (48) = .77, p < .01, între venitul familial lunar şi cheltuielile lunare pe produse cosmetice. Ambele corelații ce implică venitul lunar sunt semnificative, ceea ce ne face să credem că variabila respectivă influențează relația dintre managementul impresiei şi cheltuielile amintite. Venitul nu ar fi influențat relația dintre ele dacă nici una dintre relațiile venitului cu cele două variabile nu era semnificativă.

Pentru a vedea care este relația dintre managementul impresiei şi cheltuielile destinate produselor cosmetice, eliminând influența venitului familial lunar asupra celor două, vom calcula coeficientul de corelație parțială. Acesta are mai multe simboluri, însă noi îl vom nota cu *pr*, respectând recomandările „bibliei psihologilor" (APA, 2001). Rezultatul obținut indică pr (47) = .55, p < .01, valoare mai mică decât cea exprimată în relația directă dintre cele două variabile [r(48) = .70, p < .01]. Astfel, ținând constantă influența venitului lunar asupra relației dintre cele două, se constată, în continuare, o relație pozitivă semnificativă între preocuparea subiecților pentru managementul impresiei şi cheltuielile efectuate pentru achiziționarea de produse cosmetice. De altfel, expresiile de genul „eliminând influența lui A" sau „ținând constantă influența lui A" sunt echivalente şi specifice în formularea interpretării coeficientului de corelație parțială.

Coeficientul de corelație parțială poate fi calculat pentru mai mult de o singură variabilă moderatoare. Spre exemplu, putem calcula gradul de asociere dintre managementul impresiei şi consumul de produse cosmetice eliminând influența exercitată atât de venitul familial lunar, cât şi de sexul persoanei chestionate. Cu fiecare variabilă ținută sub control statistic se pierde câte un grad de libertate. Dacă în cazul lui r, gradele de libertate erau stabilite după formula N-2, în cazul corelației parțiale, gradele de libertate sunt obținute pe baza formulei N-2-k, unde k reprezintă numărul de variabile ținute constante. Astfel, în cazul unui pr de ordinul I (cu o singură variabilă moderatoare) vom avea N-3, în cazul unui pr de ordinul II (cu două variabile moderatoare) vom avea N-4 ş.a.m.d. În exemplul dat, valoarea 47 a gradelor de libertate a fost obținută scăzând din totalul de 50 de subiecți testați, valoarea 3, deoarece exista o singură variabilă ce trebuia ținută constantă – venitul familial lunar.

Nu toate datele sunt exprimate prin scale numerice pentru a se putea calcula r. Destul de des variabilele sunt analizate pe baza frecvenței de apariție a răspunsurilor şi ceva mai rar, pe baza ierarhiei răspunsurilor. Au fost dezvoltate o

serie de tehnici, precum C al lui Pearson sau V al lui Kramer care au la bază calcularea lui χ^2. Alegerea coeficienților de asociere în funcție de tipul de date culese este prezentată în tabelul 1.2, din capitolul destinat statisticii descriptive.[1]

3.2. Tehnici de comparație între grupuri

Ipotezele nu se rezumă doar la posibilitatea de a stabili gradul de asociere dintre variabile. De multe ori interesul cade asupra stabilirii unor diferențe semnificative între grupe. Întrebarea „Poate terapia prin râs reduce simptomele depresive?" poate fi analizată prin compararea diferențelor în ce priveşte nivelul depresiei pe două grupe: un grup experimental, care urmează o terapie prin râs şi un grup de control, asupra căruia cercetătorul nu intervine. Evaluarea diferențelor dintre grupe este dependentă de modalitatea de măsurare a acestora. Cel mai des întâlnim, fie variabile de tip nominal, ale căror răspunsuri sunt analizate pe baza diferențelor de frecvență (χ^2), fie variabile de tip numeric, analizate pe baza diferențelor dintre medii (testele t şi z).

3.2.1. Tehnicile χ^2

Distribuția χ^2 este frecvent întâlnită în cazul datelor de tip nominal, fie sub forma unui test χ^2 al gradului de potrivire (în engleză „goodness of fit"), fie sub forma unui test χ^2 al gradului de omogenitate (în engleză „homogenity test"). O situație tipică pentru primul caz este de a stabili dacă există diferențe semnificative între răspunsurile la o întrebare de genul:

„Dacă duminica următoare ar avea loc alegeri prezidențiale, v-ați prezenta la vot?"
a) Da b) Nu c) Nu ştiu / Nu răspund

Pe baza simplei analize de frecvențe, rezultatele colectate sunt sintetizate astfel:

Tabelul 3.1. Frecvența răspunsurilor celor 70 de subiecți chestionați

„Dacă duminica următoare ar avea loc alegeri prezidențiale, v-ați prezenta la vot?"		
DA	NU	NU ŞTIU – NU RĂSPUND
35	20	15

[1] Pentru detalii, vezi Clocotici şi Stan (2000), Hinkle şi colab., (1994); Rotariu şi colab., (1999); Sava (2003) sau Yaffee (1996).

În schimb, testul χ^2 al gradului de omogenitate se foloseşte atunci când dorim să vedem diferenţe (asocieri) între categoriile de răspuns de la două întrebări.

„Dacă duminica următoare ar avea loc alegeri prezidenţiale, v-aţi prezenta la vot?"

a) Da b) Nu C) Nu ştiu / Nu răspund

„Aţi votat la ultimele alegeri prezidenţiale?"

a) Da b) Nu

În acest caz, tehnica χ^2 al gradului de omogenitate va fi utilizată după sintetizarea prealabilă a răspunsurilor la cele două întrebări, într-un singur tabel:

Tabelul 3.2. Frecvenţa răspunsurilor celor 70 de subiecţi chestionaţi

„Dacă duminica următoare ar avea loc alegeri prezidenţiale, v-aţi prezenta la vot?"				
Aţi votat la ultimele alegeri prezidenţiale?		Da	Nu	Nu ştiu – Nu răspund
	Da	24	7	5
	Nu	11	13	10

Ipoteza nulă a ambelor variante ale tehnicii χ^2 stipulează că nu există diferenţe semnificative între frecvenţa răspunsurilor observate şi cea a răspunsurilor aşteptate în diferite categorii (Hinkle, Wiersma şi Jurs, 1994). Exemplul ce însoţeşte explicarea celor două tehnici χ^2 este o adaptare ipotetică a celui de-al doilea experiment realizat de Cialdini şi colab. (1976/1999):

Studenţilor li s-au adresat prin telefon şase întrebări generale ale căror răspunsuri erau greu de intuit exact, cum ar fi „ce procent de studenţi dintre colegii tăi lucrează?". După terminarea celor şase întrebări, a intervenit manipularea experimentală, subiecţii fiind împărţiţi aleatoriu în trei grupe. Celor din prima grupă li s-a dat un feedback pozitiv, spunând că au răspuns mult mai bine decât media, cinci din cele şase răspunsuri fiind corecte. Celor din grupa a doua li s-a oferit un feedback neutru, răspunsurile lor fiind medii, deoarece au răspuns corect la trei întrebări din şase. În sfârşit, cei din al treilea grup au primit un feedback negativ, experimentatorii sugerând că au răspuns mai slab decât media, un singur răspuns fiind corect. După această manipulare li s-a pus o singură întrebare suplimentară, legată de rezultatul echipei de fotbal american a universităţii respective la unul dintre ultimele două meciuri susţinute. Unii au fost întrebaţi de rezultatul meciului pierdut de echipa universităţii, iar alţii de rezultatul meciului câştigat. Răspunsurile la întrebare au fost codate în una din cele două categorii: „noi" sau „non-noi". În categoria „noi" au fost incluse răspunsuri de genul „I-am bătut", „Noi am pierdut", „Am câştigat cu 4-2", „Am fost plecat din oraş şi nu ştiu ce au făcut ai noştri", iar în categoria „non-noi" s-au regăsit răspunsuri precum „Au pierdut", „Cred că au câştigat", „Nu ştiu ce-au făcut".

Studiul avea două ipoteze: (1) răspunsurile de tip „noi" apar mai frecvent în situaţia de victorie a echipei de fotbal, decât în cea de înfrângere; (2) acest efect va fi intensificat în

cazul subiecţilor cărora li s-a dat anterior un feedback negativ. Ambele ipoteze pornesc de la premisa că o persoană tinde să se asocieze cu evenimente publice pozitive, respectiv să se distanţeze de evenimente publice negative, efectul fiind potenţat mai ales pentru subiecţii care au suferit eşecuri personale.

Prima ipoteză necesită un test χ^2 al gradului de potrivire, în timp ce a doua ipoteză necesită un test χ^2 al gradului de omogenitate. În continuare vom detalia asemănările şi diferenţierile dintre cele două forme χ^2.

Ambele forme pornesc de la ideea comparării frecvenţelor observate cu cele expectate (aşteptate). De exemplu, aruncând o monedă de 100 de ori, observăm că stema a ieşit de 54 de ori, iar banul de 46. Acestea sunt frecvenţele observate. Cele aşteptate (teoretice) sunt cele stabilite pe bază de probabilităţi. Astfel, ţinând cont că sunt doar două feţe ale monedei şi au fost 100 de aruncări, ne-am aştepta ca fiecare parte să apară de 50 de ori. Diferenţele dintre valorile frecvenţelor observate şi cele ale valorilor aşteptate sunt luate în considerare în calcularea lui χ^2.

Formula de calcul a lui χ^2 este:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \qquad (3.1)$$

unde O simbolizează frecvenţele observate, iar E frecvenţele aşteptate.

În orice tehnică de comparaţie a datelor va trebui să răspundem la următoarele întrebări:

1) Cum identificăm situaţia de comparaţie, de testare a diferenţelor?
2) Ce fel de grupe sau categorii de răspunsuri sunt comparate?
3) Ce fel de date au fost colectate?
4) Condiţiile aplicării tehnicii respective sunt îndeplinite?
5) Ce fel de ipoteză trebuie testată?
6) Cât de intensă este diferenţa dintre grupe sau dintre categoriile de răspunsuri?

(1) Cum identificăm situaţia de comparaţie, de testare a diferenţelor?

Majoritatea studiilor din psihologie au la bază teste de stabilire a semnificaţiei diferenţelor dintre grupe sau categorii de răspunsuri. Pentru a identifica aceste studii este suficient să analizăm modul de formulare a ipotezelor. În exemplul dat vom regăsi expresia: „răspunsurile de tip <<noi>> apar mai frecvent în situaţii de victorie, decât în cazuri de înfrângere". Alte sintagme care indică necesitatea comparaţiei pot fi: „grupul A are rezultate mai mari decât grupul B", „tratamentul A este mai bun decât tratamentul B" sau „ambele tratamente sunt mai bune decât absenţa oricărei intervenţii". Toate aceste formulări au ca ipoteză nulă uzuală absenţa diferenţelor dintre medii sau frecvenţe. O situaţie aparte de comparare a grupurilor este întâlnită în cazul testului χ^2 al gradului de omogenitate

(cea de-a doua ipoteză din exemplu). În acest caz, premisa ipotezei nule este că frecvența de apariție a răspunsurilor „noi" în situații de victorie sau de înfrângere nu diferă în funcție de tipul de feedback primit de subiecți.

(2) Ce fel de grupe sau categorii de răspunsuri sunt comparate?

Vom amâna până la prezentarea testelor t şi z discuția despre această problemă. Pentru moment vom preciza doar că grupele comparate pot fi independente sau perechi. În cazul lui χ^2 răspunsurile subiecților trebuie să fie independente. O încălcare a acestei condiții se întâlneşte în testarea repetată a aceluiaşi grup de subiecți, deoarece în cazul măsurătorilor repetate, rezultatele exprimate anterior sau ulterior de un subiect nu sunt independente complet. În asemenea situații trebuie aplicate alte tehnici statistice precum testul χ^2 McNamar sau χ^2 Stuart-Maxwell (Hinkle, Wiersma şi Jurs, 1994).

(3) Ce fel de date au fost colectate?

Variabilele independente din cercetare pot fi considerate, fie de tip ordinal, cu trei categorii de răspunsuri: feedback pozitiv, feedback neutru, respectiv feedback negativ, fie de tip nominal, cu două categorii de răspunsuri – meci câştigat, respectiv meci pierdut. Variabila dependentă a cercetării este tot de tip nominal, categoriile de răspunsuri posibile fiind „noi" şi „non-noi". Acestea pot fi reflectate şi dihotomic sub forma 1-răspunsuri de tip „noi" şi 0-răspunsuri de tip „non-noi". În cazul comparațiilor dintre grupe, o importanță deosebită o are modul de măsurare a variabilei dependente (criteriu). În cazul unor date de tip nominal, ca în situația de față, se va calcula χ^2. În schimb, în cazul unor date numerice, se vor utiliza, cu mici excepții, tehnicile t sau z.

Tabelul 3.3. Procentajul subiecților care au dat răspunsuri de tip „noi"

	Tip de feedback			
Tip de meci	Pozitiv %	Neutru %	Negativ %	Media %
Meci câştigat	28 (7/25)	28 (7/25)	64 (16/25)	41 (30/75)
Meci pierdut	36 (9/25)	32 (8/25)	16 (4/25)	28 (21/75)

(4) Condițiile aplicării tehnicii respective au fost îndeplinite?

Pentru aplicarea testului χ^2 există câteva condiții şi restricții de utilizare:

- observațiile trebuie să fie independente, fiind excluse design-urile cu măsurători repetate; în plus, categoriile de răspunsuri trebuie să se excludă reciproc, ceea ce presupune că întrebările care permit răspunsuri multiple nu pot fi analizate prin χ^2.
- trebuie să fie de cel puțin 20 de frecvențe observate; mai mult, frecvența aşteptată pentru oricare căsuță (celulă) a tabelului de contingență nu trebuie să fie mai mică de 1; similar, un procent de peste 20% din căsuțe

care au frecvențe teoretice mai mici de 5 duce la o valoare χ^2 discutabilă; o tactică posibilă în asemenea situații este reunirea unor căsuțe adiacente.

- în cazul unor tabele de contingență de tip 2 x 2, este necesară aplicarea unei corecții Yates ce presupune creşterea cu 0,5 a diferenței dintre frecvențele observate şi cele teoretice.
- în calculul χ^2 se ține cont de frecvențele şi nu de procentajele constatate, deoarece cele din urmă duc la calcularea χ^2 ca şi cum ar fi fost obținute 100 de răspunsuri indiferent dacă au fost colectate 5000, 500 sau 50 de observații.

În exemplul nostru, toate condițiile sunt îndeplinite, existând 51 de răspunsuri de tip „noi" într-un design experimental de tip 3 x 2.

(5)Ce fel de ipoteze au fost testate?

În general rămâne valabilă distincția dintre ipotezele unidirecționale şi cele bidirecționale. Ceea ce se schimbă ține de sintagmele utilizate. Astfel, în compararea diferențelor dintre grupe, ipotezele unidirecționale pot fi de genul „fetele au un nivel de empatie mai ridicat decât cel al băieților", „metoda A duce la un număr de greşeli mai mic comparativ cu metoda B" etc. Ipotezele bidirecționale nu specifică direcția diferențelor dintre grupe, fiind mai degrabă formulări vagi precum „există diferențe între fete şi băieți în ce priveşte nivelul de empatie" sau „metodele A şi B sunt diferite în ce priveşte numărul de greşeli comise".

Totuşi, în cazul lui χ^2, distincția dintre ipotezele unidirecționale şi cele bidirecționale poate fi aplicată doar în cazul unui test cu un singur grad de libertate. Prima ipoteză poate fi considerată ca fiind de tip unidirecțional, deoarece χ^2 are un singur grad de libertate, iar cercetătorii anticipează că răspunsurile de tip „noi" apar mai frecvent în situația de victorie a echipei de fotbal, decât în cea de înfrângere. În schimb, în cazul celei de-a doua ipoteze, χ^2 poate fi doar bidirecțional, deoarece χ^2 are două grade de libertate, fiind deja un test omnibuz. Acesta arată dacă există diferențe între frecvențe, însă nu arată unde se află acele diferențe. Astfel, în cazul unui χ^2 semnificativ vom afla că tipul de feedback primit influențează diferit nivelul răspunsurilor de tip „noi" în funcție de rezultatul meciului echipei universității. El nu ne va arăta însă nici ce fel de feedback influențează răspunsurile subiecților, nici dacă aceste răspunsuri vor creşte sau descreşte în frecvență. Tot ce vom afla este dacă frecvența se modifică sau nu. Pentru a suplini limitele χ^2, există modalități specifice de lucru, precum analiza reziduurilor (Hinkle, Wiersma şi Jurs, 1994).

Tabelul 3.4. Frecvența observată şi aşteptată privind subiecții care au dat răspunsuri de tip „noi" în funcție de tipul de meci chestionat.

Tip de meci	Răspunsuri de tip „noi" observate	Răspunsuri de tip „noi" aşteptate
Meci câştigat	30	25,5
Meci pierdut	21	25,5
Total subiecți	51	51

Dacă nu există diferențe între cele două categorii de subiecți chestionați, ne aşteptăm ca în fiecare situație să existe 25,5 răspunsuri de tip „noi". Se observă că frecvența observată este diferită de valorile aşteptate. Se pune întrebarea simplă: este această diferență întâmplătoare sau diferența este suficient de mare pentru ca probabilitatea de a obține astfel de date din întâmplare să fie mai mică de .05? Cu alte cuvinte, va trebui să testăm în ce măsură diferența este suficient de mare pentru a sprijini afirmația că răspunsurile de tip „noi" apar mai frecvent în situație de victorie a echipei de fotbal aparținătoare decât în cazul înfrângerii acesteia.

Tehnic, pentru a calcula χ^2 în SPSS avem două direcții posibile, în funcție de tipul testului (χ^2 al gradului de potrivire, respectiv χ^2 al gradului de omogenitate). Pentru a testa această primă ipoteză ne aflăm în situația unui χ^2 al gradului de potrivire. Metoda poate fi selectată urmând calea „Analyze"-„Nonparametric tests"-„Chi-square". Opțiunea standard din SPSS este de a calcula frecvența teoretică împărțind totalul frecvențelor la numărul de categorii. Această modalitate este valabilă şi pentru problema de față (51:2). Totuşi, există şi o altă cale de calculare a frecvențelor aşteptate, pe baza unor informații a priori. Spre exemplu, dacă frecvența de apariție a nevrozei într-o populație este de 20%, iar noi suntem interesați să vedem dacă frecvența de apariție în grupul testat este diferită față de populația din care a fost extras eşantionul, atunci frecvența teoretică va fi de 20% din numărul subiecților testați.

În cazul primei ipoteze vom obține un $\chi^2(1) = 1.59$, $p < .10$, test unilateral, un rezultat nesemnificativ statistic. Prin urmare, nu putem afirma că răspunsurile de tip „noi" apar mai frecvent în situații de victorie a echipei de apartenență decât în situații de înfrângere. Totuşi, valoarea lui p apropiată de pragul maxim admis de .05, ne duce cu gândul că am putea asista la o schimbare a concluziei, dacă vom creşte puterea statistică a cercetării.

χ^2 al gradului de omogenitate poate fi calculat în SPSS utilizând o cale diferită, ceva mai complexă: „Analyze"-„Descriptive Statistics"-„Crosstabs", după care se alege butonul „Statistics", iar din fereastra deschisă vom selecta „Chi-square". În plus, revenind în Crosstabs, se poate selecta şi butonul „Cells" pentru a alege reziduurile standardizate ajustate. Asupra rolului acestora vom reveni curând.

Testarea celei de a doua ipoteze poate fi verificată utilizând baza de date *„hi patrat.sav"*, după ce în prealabil am ponderat variabila „frecv" utilizând calea „Data"-„Weight cases". În urma demersului s-a obținut un $\chi^2(2) = 6,12$, $p < .05$. Valoarea χ^2 este semnificativă statistic, sprijinind ipoteza cercetării: frecvența răspunsurilor de tip „noi" în caz de victorie sau înfrângere a echipei sportive de apartenență este influențată diferit în funcție de tipul de feedback oferit subiecților. Mai mult, din consultarea figurii 3.7.se observă tendința persoanelor care au primit feedback negativ să utilizeze mai multe răspunsuri de tip „noi" în cazul în care echipa universității a câştigat.

Frecvența aşteptată pentru fiecare căsuță este rezultatul produsului dintre totalurile marginale corespunzătoare căsuței respective împărțit la totalul frecvențelor observate. Spre exemplu, frecvența aşteptată a răspunsurilor de tip „noi" în cazul persoanelor care au primit feedback neutru şi care ofereau rezultatul meciului pierdut de echipa universității este de 6,2 [(21x15)/51].

			Tipul de feedback			
			feedback pozitiv	feedback neutru	feedback negativ	Total
Rezultatul meciului	victorie	Frecventa observata	7	7	16	30
		Frecventa asteptata	9,4	8,8	11,8	
		Adjusted Residual	-1,5	-1,1	2,5	
	infrangere	Frecventa observata	9	8	4	21
		Frecventa asteptata	6,6	6,2	8,2	
		Adjusted Residual	1,5	1,1	-2,5	
Total		Frecventa observata	16	15	20	51

Figura 3.7. *Output SPSS cu frecvențele observate, cele aşteptate şi reziduurile standardizate ajustate privind subiecții care au dat răspunsuri de tip „noi" în funcție de tipul de meci chestionat şi tipul de feedback primit*
SPSS este o marcă înregistrată.

Testul χ^2 este un test omnibuz, adică arată doar la nivel general că tipul de feedback primit afectează în mod diferit frecvența răspunsurilor de tip „noi" în funcție de tipul de meci vizat. Pe baza frecvențelor observate putem constata unde se află aceste diferențe, însă o asemenea perspectivă este foarte subiectivă.

De aceea, în cazurile unui χ^2 semnificativ statistic se recomandă calcularea reziduurilor standardizate ajustate pentru fiecare celulă în parte, pentru a observa unde se situează acele diferențe. Valorile reziduurilor standardizate ajustate cuprinse în intervalul [-2, 2] nu ne interesează. În schimb, valorile care nu se regăsesc în acest interval sunt considerate ca indicatori ai unei diferențe semnificative între frecvențele observate şi cele aşteptate. Valoarea absolută de 2 este stabilită pe baza unei distribuții z, a cărei valoare corespunzătoare pentru un p de .05 este 1,96.

În figura 3.7 atrag atenția două reziduuri standardizate ajustate, ambele fiind legate de oferirea unui feedback negativ subiecților. Astfel, se constată o frecvență observată semnificativ mai mare decât cea aşteptată în situația unui feedback negativ oferit persoanei şi a unui rezultat pozitiv al echipei universității. Inversul acestei situații apare în situația unui feedback negativ şi a unei înfrângeri a echipei universității.

Ce înseamnă acest lucru? Dacă χ^2 ne arăta că tipul de feedback afectează în mod diferit referința directă la grupul de apartenență în funcție de tipul de meci vizat, analiza reziduurilor ne arată că afirmația de mai sus este valabilă în cazul unui feedback negativ. Astfel, subiecții cărora li s-a spus că au obținut performanțe

sub medie la întrebările inițiale tind să folosească mai des, în caz de victorie a echipei, sintagme de tipul „noi am bătut", „echipa noastră a câştigat" etc. Similar, subiecții care au primit un feedback negativ tind să folosească în caz de înfrângere mai puține sintagme de identificare cu echipa universității.

Dacă în locul răspunsurilor de tip „noi" ar fi fost contorizate frecvențele celor de tip „non-noi", am fi obținut o situație în oglindă. Adică, persoanele care au primit un feedback negativ tind să folosească mai rar, în cazul victoriei echipei de apartenență, expresii de genul „au bătut" sau „ei au câştigat". Astfel, se poate concluziona că subiecții al căror prestigiu personal a fost afectat tind să utilizeze pronumele personal „noi" sau expresii care îi includ implicit pentru a se asocia cu situații de succes ale grupului de apartenență.

Înainte de a discuta mărimea efectului pentru tehnicile χ^2, vom prezenta, pe scurt, modul de stabilire a gradelor de libertate. În cazul unui χ^2 al gradului de potrivire, gradele de libertate se stabilesc uşor, după formula c-1, unde c reprezintă numărul de categorii de răspuns posibile. În prima ipoteză întâlnim două categorii: meciuri câştigate sau meciuri pierdute de echipa de fotbal a universității. Prin urmare există un singur grad de libertate. În schimb, în cazul unui χ^2 al gradului de omogenitate, specific tabelelor de contingență, formula devine (c-1)(l-1), unde c înseamnă numărul de coloane (categorii ale primei variabile), iar l înseamnă numărul de linii (categorii ale celei de-a doua variabile). În cea de-a doua ipoteză vom avea 2 grade de libertate (3-1)(2-1), deoarece tipul de feedback are trei categorii de răspunsuri – feedback pozitiv, neutru sau negativ, iar tipul de meci rămâne cu cele două categorii – meciuri câştigate, respectiv pierdute.

(6) Cât de mare este diferența dintre grupe sau categoriile de răspunsuri?

Un rezultat semnificativ statistic nu înseamnă decât că grupele comparate diferă între ele. El nu înseamnă neapărat că diferența este mare. Pentru a vedea cât de mare este diferența dintre grupe sau categorii de răspunsuri vom apela la indicatorii de apreciere a mărimii efectului.

Din păcate, majoritatea indicatorilor pornesc de la ideea unei variabile criteriu (dependente) măsurate prin scale numerice, fapt ce afectează calcularea mărimii efectului în cazul testelor χ^2. Totuşi, printr-un artificiu prezentat în Rosenthal, Rosnow şi Rubin (2000a), putem estima mărimea efectului în mod clasic doar în cazul testelor χ^2 cu un singur grad de libertate. Coeficientul de determinare poate fi extras pe baza formulei:

$$r^2 = \frac{\chi^2(1)}{N} \qquad (3.2)$$

unde $\chi^2(1)$ reprezintă valoarea testului de comparație a diferenței dintre frecvențe, cifra 1 dintre paranteze ne aminteşte că formula este valabilă doar dacă testul are un singur grad de libertate, iar N este numărul total de frecvențe observate.

De altfel, N reprezintă şi valoarea maximă a lui χ^2 în situația respectivă. În cazul primei ipoteze poate fi calculat coeficientul de determinare, deoarece există un singur grad de libertate - $\chi^2(1) = 1,93$, $p > .10$. Această valoare este comparată cu χ^2 maxim pentru situația respectivă, care este 51, deoarece totalul răspunsurilor de tip „noi" a fost 51. În consecință, r^2 este .037 sau 3,7%. Dacă vom apela la criteriile de comparație prezentate în tabelul 2.2 din capitolul destinat principiilor testării ipotezelor, vom constata că există un efect relativ redus al rezultatului meciului echipei universității asupra tendinței subiecților de a formula răspunsuri de tip „noi". Mai mult, dacă vom ține cont că r^2 nu este decât pătratul coeficientului de corelație, vom putea observa că unui $\chi^2(1) = 1,93$, $p > .10$ îi corespunde o corelație r de .19 între cele două variabile.

Atât în cazul unui $\chi^2(1)$, cât şi în cazul unui χ^2 cu mai mult de un grad de libertate, poate fi estimat un alt coeficient al mărimii efectului, notat *w*. Acest indicator poate fi calculat prin intermediul software-ului GPOWER (Faul şi Erdfelder, 1992), disponibil gratuit de pe internet.

Mărimea efectului în cazul primei ipoteze $\chi^2(1) = 1,93$, $p > .10$ este .18, o valoare care indică acelaşi efect slab al rezultatului partidei de fotbal asupra utilizării răspunsurilor de tip „noi". Pentru a avea o impresie generală, un w de .10 semnifică un efect slab, un w de .30 semnalează un efect de intensitate medie, iar un w de .50 sau mai mare indică un efect puternic. Pe baza mărimii efectului vom constata că puterea statistică de testare a primei ipoteze nu era mai mare de .25, ceea ce înseamnă de trei ori mai puține şanse de a observa acest efect mic, dacă el există într-adevăr, decât de a comite eroarea de tip II. Pentru a atinge un nivel optim al puterii statistice de .80, necesar testării ipotezei, am fi avut nevoie de mai mulți subiecți. Mai exact, am fi avut nevoie de 122 de răspunsuri, cu 71 mai multe decât cele 51 de răspunsuri observate. În consecință, datorită puterii reduse, rezultatul nesemnificativ statistic are o explicație alternativă ipotezei nule. Diferențele pot exista în realitate, însă nu au fost identificate datorită unui risc crescut de a comite eroarea de tip II. În cazul unei puteri statistice de .25, probabilitatea de eroare este de 75% (1 – puterea statistică).

Utilizând acelaşi program statistic, vom constata un w de .38 în cazul celei de-a doua ipoteze, ceea ce indică o schimbare de magnitudine medie în utilizarea răspunsurilor de tip „noi" în situații de câştig sau pierdere, în funcție de tipul de feedback personal primit. În cea de-a doua ipoteză, $\chi^2(1) = 6,12$, $p < .05$, puterea statistică era de .72, un nivel destul de apropiat de cel optim. De altfel, pentru a atinge valoarea optimă de .80 a puterii statistice, am fi avut nevoie de 61 de răspunsuri de tip „noi", cu 10 mai multe decât datele colectate. În consecință, rezultatul semnificativ statistic corelat cu o mărime medie a efectului celor două VI asupra VD, sprijină clar ideea că subiecții, al căror prestigiu personal a fost afectat, tind să utilizeze mai frecvent termeni de referință de genul pronumelui „noi" pentru a se asocia cu situații de succes ale grupului de apartenență, ca o măsură de auto-înălțare.

În sfârşit, o altă modalitate de calculare a mărimii efectului în cazul χ^2 constă în utilizarea indicatorilor de asociere obținuți pe baza valorii χ^2, precum V al lui Cramer. Acesta este un coeficient de corelație non-direcțional ce variază între 0 şi 1, unde 0 semnifică absența efectului, iar 1 un efect extrem de puternic.

De altfel, întregul demers χ^2 prezentat anterior poate fi interpretat şi sub forma asocierii dintre două variabile nominale. Într-o asemenea situație, ipoteza nulă indică absența relației dintre două variabile, în timp ce un χ^2 semnificativ statistic sprijină ipoteza asocierii dintre două variabile. De aceea, într-un fel, χ^2 constituie un liant între tehnicile de asociere şi cele de comparație.

3.2.2. Testele t şi z de comparare a diferențelor dintre două medii

Dacă variabila dependentă nu este măsurată prin scale nominale pentru a fi supusă analizei de frecvență, ci este evaluată prin scale numerice, este nevoie de compararea mediilor pentru a testa ipoteza existenței diferențelor dintre două populații pornind de la compararea a două grupe. De altfel, cele mai multe dintre cercetările studenților la psihologie necesită aplicarea testelor t sau z. Există trei tipuri principale de tehnici t şi z, însă SPSS-ul permite doar calcularea testelor t. Valorile z pot fi totuşi deduse din rezultatele oferite.

Cele trei tipuri de teste t şi z sunt:

- tehnici care privesc diferența dintre media unui eşantion şi cea a populației din care acesta a fost extras; titulatura tehnicilor este testul t pentru un eşantion, respectiv testul z pentru un eşantion; pentru a efectua, în SPSS, testul t pentru un eşantion vom urma calea „Analyze"-„Compare Means"-„One Sample T Test".
- tehnici care privesc diferența dintre mediile a două eşantioane independente; ele sunt denumite testul t pentru eşantioane independente, prescurtat t independent, respectiv testul z pentru eşantioane independente sau z independent; pentru a calcula t independent în SPSS vom urma calea „Analyze"-„Compare Means"-„Independent Samples T Test".
- tehnici care privesc diferența dintre mediile a două eşantioane perechi (corelate, dependente); în asemenea situații vorbim de testul t pentru eşantioane perechi, prescurtat t dependent, respectiv testul z pentru eşantioane perechi, ori z dependent; în SPSS, t dependent poate fi obținut pe calea „Analyze"-„Compare Means"-„Paired Samples T Test".

Ce sunt eşantioanele independente şi eşantioanele perechi?

În ce priveşte diferențierea dintre testele t şi z independente şi dependente, ea se face în funcție de natura grupelor comparate. Astfel apar două situații:

(1) dacă cele două grupe comparate au fost alese la întâmplare, pe baza situației lor naturale, ca în cazul a două clase paralele de elevi sau a două grupe de vârstă diferite, se utilizează testele pentru eşantioane independente.

(2) dacă cele două grupe sunt în relație una cu cealaltă prin intervenția experimentatorului sau prin corespondența naturală a elementelor din cele două grupe, vom utiliza testele pentru eşantioane perechi (corelate). Există trei situații în care vorbim de eşantioane dependente (perechi):

a) *situația măsurătorilor repetate*
O grupă de subiecți este testată de două ori asupra aceleiaşi variabile, cercetătorul fiind interesat de modificările apărute între cele două testări, pre-test şi post-test. Testând aceleaşi persoane, se ajunge la perechi de rezultate, de pildă scorurile obținute de „X" la începutul şi la finalul unei intervenții terapeutice.

b) *situația perechilor naturale*
Cele două grupe sunt formate după una sau mai multe caracteristici prealabile, care există în mod natural. În cazul studiilor privind importanța ordinii la naştere se pot crea două grupe de subiecți, dintre care într-o grupă să fie incluşi primii născuți, iar în cealaltă grupă mezinii, cu condiția ca fiecare dintre aceştia să aibă un corespondent (un frate) în prima grupă. Se ajunge, din nou, la perechi de rezultate, de pildă scorurile obținute de „Popescu A.", primul născut, distribuit în grupa A şi scorurile obținute de „Popescu Z.", fratele mai mic al primului, distribuit în grupa B.

c) *situația perechilor artificiale*
Elementele celor două grupe sunt puse în corespondență prin intervenția experimentatorului. Această tehnică este utilizată cu scopul de a egaliza grupele ce urmează a fi comparate, pentru a elimina influența unor variabile terțe. De exemplu, dacă vrem să vedem în ce măsură metoda de predare A este mai bună decât metoda B, poate ar fi bine ca grupele ce urmează a fi învățate prin metode diferite să fie echilibrate, adică alcătuite din elevi de calitate similară. De exemplu, ele pot fi egalizate printr-un test inițial de cunoştințe. Subiecții care au obținut primele două locuri vor fi distribuiți în grupe diferite şi se va realiza, ulterior, o corespondență între rezultatele lor. Algoritmul de lucru se va păstra, până la finalizarea distribuirii în grupele de lucru nou formate a subiecților testați prealabil.

Se observă că eşantioanele perechi vor avea întotdeauna un număr de subiecți egal, pe când eşantioanele independente au o libertate mai mare, cele două grupe comparate putând avea sau nu un număr egal de subiecți.

Iată câteva exemple ilustrative de studii care necesită aplicarea testelor t:

Testul t pentru două eşantioane independente

Un cercetător crede că strategiile sintetice conduc mai repede la deprinderea cititului decât strategiile analitice. Pentru a testa ipoteza el selectează două clase paralele de elevi aflaţi în clasa I, fiecare dintre ele având câte 20 de elevi. Primul grup a fost învăţat să citească analitic (începând cu literele, trecând apoi la cuvinte), în timp ce al doilea grup a învăţat să citească sintetic (pornind de la asocierea cuvintelor cu imagini, pentru a deprinde ulterior literele). Variabila dependentă a constat într-o probă de citire contra cronometru, obţinându-se un indicator pe baza formulei: numărul de cuvinte citite într-un minut / numărul de erori.

Testul t pentru două eşantioane perechi – cazul măsurătorilor repetate

Obiectivul cercetării este de a studia influenţa metodelor de relaxare asupra reducerii conduitelor agresive la adolescenţi. O grupă de 15 subiecţi cu tulburări comportamentale este testată iniţial (pre-test) în ceea ce priveşte nivelul agresivităţii, printr-un inventar de personalitate. După testare, cei 15 urmează un program de relaxare, constituit din cinci şedinţe de relaxare săptămânale, fiecare durând o oră. După o perioadă de două luni subiecţii sunt retestaţi (post-test) pentru a observa dacă nivelul lor de agresivitate a scăzut în urma practicării tehnicilor de relaxare învăţate.

Testul t pentru două eşantioane perechi – cazul perechilor naturale

Un cercetător norocos găseşte 20 de perechi de gemeni univitelini, dar crescuţi în familii diferite, fiind despărţiţi de la naştere. Mai mult, diferenţa era de ordin calitativ, unii crescând în familii „normale", iar ceilalţi în familii cu „probleme". Obiectivul studiului său este de a stabili în ce măsură agresivitatea este învăţată social, cu alte cuvinte în ce măsură există diferenţe semnificative în manifestările agresive ale gemenilor din cele două grupe. Agresivitatea a fost codată pe baza unui protocol de observaţie, scorurile variind între 4 şi 20.

Testul t pentru două eşantioane perechi – cazul perechilor artificiale

Obiectivul cercetării este de a studia efectul atmosferei de lucru asupra rezolvării problemelor de matematică. Cercetătorul decide să împartă subiecţii în cele două grupe de lucru pe baza unui criteriu prestabilit. Astfel, sunt testate iniţial cunoştinţele de matematică ale elevilor, iar pe baza rezultatelor obţinute are loc distribuirea elevilor în cele două grupe, pentru a controla (elimina) influenţa acestora asupra relaţiei dintre variabilele studiate. Pentru a echivala cele două grupe, cercetătorul ia primii doi subiecţi, care au obţinut rezultatele cele mai bune la testul iniţial, şi îi distribuie în grupe diferite. Apoi trece la următorii doi subiecţi pe care îi distribuie, de asemenea, în grupe diferite. Se continuă algoritmul până sunt divizaţi toţi subiecţii. Apoi, una dintre cele două grupe nou-formate va rezolva probleme de matematică într-o atmosferă calmă, relaxantă, în timp ce al doilea grup va rezolva probleme de matematică într-o atmosferă tensionată, conflictuală. Variabila dependentă este performanţa obţinută de fiecare elev, exprimată în numărul de probleme rezolvate corect, maximum fiind 10.

Testul t pentru un singur eşantion

Un test de empatie a fost administrat pe o populaţie de elevi proaspăt intraţi la liceu (N = 1250), obţinându-se un scor mediu de 76. La una din clasele testate (n = 28) la aceeaşi probă s-a obţinut un scor mediu de 80 şi o abatere standard de 8. Se ridică întrebarea: are această clasă un nivel mediu de empatie mai ridicat decât media populaţiei de elevi?

Dincolo de aceste diferenţe, testele de comparaţie sunt similare, cel puţin din perspectiva interpretării datelor. De aceea, vom prezenta doar două exemple, cazul unui test t independent, respectiv al unui test t dependent. De altfel, orice studiu care implică stabilirea diferenţelor dintre două medii necesită răspunsuri clare la următoarele întrebări:

1) Cum identificăm o situaţie de comparare între două grupe?
2) Ce fel de grupe sunt comparate?
3) Ce fel de date au fost colectate?
4) Condiţiile aplicării tehnicii respective sunt îndeplinite?
5) Ce fel de ipoteză trebuie testată?
6) Cât de intensă este diferenţa dintre grupe sau dintre categoriile de răspunsuri?

Problema A

Un student, cu un bogat spirit de observaţie, a remarcat că oamenii tind să se împartă în două categorii: cei care sărează mâncarea înainte de a o gusta, din obişnuinţă, respectiv cei care mai întâi gustă şi apoi îşi sărează mâncarea în funcţie de gust. Pe baza acestei categorii create ad-hoc, dorim să studiem dacă există diferenţe între cele două grupe de oameni în ce priveşte nivelul lor de impulsivitate. Această trăsătură de personalitate a fost măsurată printr-o scală specifică pe un număr de 63 de persoane, dintre care 27 intrau în prima categorie, a celor care sărează mâncarea înainte de a o gusta, iar 36 erau persoane din a doua categorie. Datele obţinute pot fi consultate în baza de date „*t independent.sav*". (o cercetare pe o temă similară a fost efectuată în realitate de McGee şi Snyder, fiind descrisă în Kerlinger şi Lee, 2000).

Problema B

Un student din anul III este interesat să stabilească dacă anul I Psihologie este mai dificil decât anul al II-lea, din perspectiva rezultatelor şcolare. Pentru aceasta, el selectează la întâmplare un grup de 28 de colegi, cărora le calculează media generală obţinută de aceştia în primii doi ani, anticipând că subiecţii selectaţi vor avea rezultate mai mari în anul II decât în anul I. Datele obţinute pot fi găsite în baza de date „*t dependent.sav*".

(1) Cum identificăm o situaţie de comparare între două grupe?

Răspunsul este foarte simplu: trebuie să existe doar două modalităţi ale VI. În prima problemă, variabila de grupare are două categorii: persoane care sărează mâncarea înainte de a o gusta şi persoane care sărează mâncarea după ce au gustat-o, după preferinţă. Dacă variabila ar fi inclus o categorie suplimentară, cea a persoanelor care nu sărează deloc mâncarea, am fi avut trei categorii VI, deci nu am fi aplicat testele t sau z. În problema B, avem tot două grupe, prima fiind constituită din rezultatele studenţilor când erau în anul I, iar cea de-a doua, din rezultatele studenţilor obţinute în anul II.

(2) Ce fel de grupe sunt comparate?

Aici trebuie să facem distincția între situația comparării mediei unui eşantion cu cea a populației (testele t şi z pentru un eşantion); situația comparării mediilor a două eşantioane independente (t independent); respectiv a comparării mediilor a două eşantioane perechi (t dependent). În problema A identificăm situația a două eşantioane independente, deoarece este vorba de persoane diferite în cele două grupe (cei care folosesc sarea înainte, respectiv cei care folosesc sarea după mâncare). În plus, între subiecții celor două grupe nu există o corespondență prealabilă naturală sau artificială. În schimb, problema B ilustrează situația a două eşantioane perechi, fiind vorba de testarea repetată a aceloraşi subiecți.

(3) Ce fel de date au fost colectate?

Pentru a aplica testele t sau z avem nevoie de compararea diferențelor dintre două medii. Cum acestea au sens numai în cazul datelor măsurate prin scale numerice, înseamnă că testele t şi z pot fi aplicate doar în cazul în care VD (criteriul) este măsurat prin scale de tip interval sau proporții. Dacă VD a fost măsurat prin scale nominale sau ordinale trebuie aplicate alte tehnici, testele t şi z nefiind adecvate. Pe lângă aceste cerințe vizând VD, reamintim că testele t şi z pot fi aplicate doar pentru a testa diferențele dintre *două* medii. În cazul comparării simultane a mai multor medii se aplică tehnicile ANOVA. În cele două probleme-exemplu se constată că datele permit aplicarea testelor t sau z, deoarece există doar două grupe în fiecare problemă, iar rezultatele VD sunt exprimate numeric (punctajul la testul de impulsivitate, respectiv media generală anuală).

(4) Condițiile aplicării tehnicii respective sunt îndeplinite?

Pentru a putea aplica în mod corect testul ales trebuie îndeplinite unele condiții statistice. O condiție valabilă pentru toate testele t şi z este aceea ca distribuția variabilei dependente (a rezultatelor) să fie normală. În cazul în care se constată devieri grave de la normalitatea distribuției, datele pot fi transformate, după indicațiile prezentate în finalul capitolului de statistică descriptivă. Dacă aceste transformări nu rezolvă problema, se renunță la aplicarea testelor t şi z, în favoarea unor tehnici non-parametrice.

În cazul testului t independent există o condiție suplimentară, aceea ca cele două grupe să aibă o dispersie omogenă (dispersii relativ egale). Condiția este analogă ideii de homoscedasticitate din cazul corelațiilor şi al regresiilor. Totuşi, dacă această condiție este încălcată, se poate calcula o variantă t ajustată, a cărei formulă este similară testelor z.

În cazul testului t dependent nu e nevoie ca cele două dispersii să fie omogene, însă există o altă condiție suplimentară, aceea a unei distribuții normale a diferenței scorurilor dintre pre-test şi post-test. Ambele exemple respectă toate condițiile de aplicare a testelor t.

Aceleaşi condiții se păstrează şi în cazul testelor z. Totuşi, *când utilizăm un test t sau un test z?* Răspunsul este mai complicat decât pare la prima vedere. În mod formal, cele două teste au distribuții statistice diferite. Totodată, pentru a aplica testul z nu trebuie să cunoaştem abaterile standard ale grupelor comparate, ci abaterile standard ale populațiilor din care sunt extrase cele două grupe. Această condiție nu poate fi îndeplinită practic decât foarte rar. Totuşi, conform teoremei limitelor centrale (Kerlinger şi Lee, 2000), în cazul unui număr de subiecți suficient de mare (cel puțin 30 de subiecți pentru fiecare grupă) distribuția rezultatelor se apropie de forma unei distribuții normale z, iar estimarea abaterii standard a populației pe baza rezultatelor obținute în cele două grupe este destul de adecvată. De aceea, se obişnuieşte să se aplice testele z când există cel puțin 30 de subiecți în fiecare grupă comparată, respectiv testele t, dacă în cel puțin unul dintre loturile comparate există mai puțin de 30 de subiecți.

Totuşi, în practică se întâlnesc des situații diferite de cele postulate în teoremă, făcând neadecvată aplicarea testelor z. În plus, în destul de multe cazuri, formula de calcul dintre t şi z nu diferă. De aceea, există tendința de a renunța la utilizarea testelor z în favoarea testelor t, ultimele putând înlocui cu succes testele z. De altfel, unele programe statistice, precum SPSS-ul, au renunțat la a prezenta testele z, diferențele dintre două medii putând fi testate pe baza testelor t.

Chiar dacă vom aplica testele z utilizând criteriul clasic al grupelor formate din peste 30 de persoane, se observă că ambele exemple necesită aplicarea testelor t, deoarece în problema A, una dintre grupe avea mai puțin de 30 de subiecți (grupul celor 27 de subiecți care sărează mâncarea înainte de a o gusta), iar problema B avea în ambele grupe câte 28 de rezultate. Testele z ar fi avut însă avantajul unor valori standard pentru a decide dacă două medii diferă semnificativ. Astfel, pentru a sprijini ipoteza că două medii diferă la un nivel semnificativ statistic, ar trebui ca valoarea calculată z să fie mai mare sau egală cu:

$z = 1{,}96$ pentru $p < .05$ sau $z = 2{,}58$ pentru $p < .01$ în cazul testelor bidirecționale;

$z = 1{,}65$ pentru $p < .05$ sau $z = 2{,}33$ pentru $p < .01$ în cazul testelor unidirecționale.

(5) Ce fel de ipoteză trebuie testată?

Înainte de rezolvarea unei probleme trebuie să stabilim tipul ipotezei de cercetare: unidirecțională (unilaterală) sau bidirecțională (bilaterală). În problema A avem o ipoteză bilaterală, deoarece sintagma utilizată „există diferențe între cele două grupe în ceea ce priveşte nivelul de impulsivitate" nu arată direcția diferenței. În schimb, problema B implică o ipoteză unilaterală: „rezultatele obținute de studenți în anul II vor fi mai mari decât cele obținute de ei în anul I". Formulările unidirecționale sunt mai precise, precum „A este mai mic decât B" sau „A este mai mare decât B", fiind specificată direcția de evoluție a diferențelor. Acest lucru nu se întâmplă în cazul unei ipoteze bilaterale de tipul „A este diferit de B".

Pe baza răspunsurilor la aceste întrebări, putem concluziona că este nevoie de un test t pentru eşantioane independente în cazul problemei A, respectiv

de un test t pentru eşantioane perechi în cazul problemei B. Vom aborda pe rând cele două probleme.

Tabelul 3.5. Mediile şi abaterile standard privind nivelul de impulsivitate din cele două grupe

	Sărează înainte de a gusta mâncarea		Sărează după ce a gustat mâncarea	
	M	AS	M	AS
Nivel de Impulsivitate	14,81	1,78	13,64	2,22

Distanța de 1,17 puncte dintre medii este suficient de mare pentru a sprijini ipoteza existenței unei diferențe în nivelul de impulsivitate între persoanele care sărează mâncarea înainte de a gusta şi cele care folosesc sarea după ce au gustat mâncarea? Constituie prima situație un semn de impulsivitate? Pentru a răspunde la aceste întrebări va trebui să respingem ipoteza nulă care neagă existența unor diferențe între cele două populații de subiecți, iar pentru aceasta vom aplica testul t pentru eşantioane independente.

		Testul Levene de egalitate a dispersiilor		Testul t de egalitate a mediilor		
		F	Sig.	t	df	Sig. (2-tailed)
Nivel de impulsivitate	Dispersie omogena	2,164	,146	2,262	61	,027
	Dispersie eterogena			2,335	60,702	,023

Figura 3.8. *Informațiile esențiale din output-ul oferit de SPSS în cazul testului t pentru eşantioane independente*
SPSS este o marcă înregistrată.

Pentru a respinge ipoteza nulă trebuie să arătăm că diferențele sunt suficient de clare încât şansele ca ele să apară din întâmplare să fie mai mici de .05. Urmând calea „Analyze"-„Compare Means"-„Independent Samples T Tests" şi stabilind grupele de comparat prin acționarea butonului „Grouping Variable" vom obține rezultatele din SPSS, prezentate în figura 3.8.

Se observă că SPSS-ul oferă două valori ale testului t, pentru fiecare situație posibilă: în cazul respectării condiției de omogenitate a dispersiei celor două grupe - $t(61) = 2.26$, $p < .05$, test bilateral, iar în cazul încălcării condiției amintite, o variantă ajustată a lui $t(60,70) = 2.33$, $p < .05$, test bilateral. Pentru a decide care dintre valori este adecvată cazului de față vom consulta valoarea testului F al lui Levene. Un F semnificativ statistic indică faptul că cele două dispersii sunt eterogene, încălcând astfel condiția de aplicare a testului t clasic. În asemenea condiții se citeşte cel de-al doilea t. Pentru problema dată, testul Levene nu este semnificativ statistic $F = 2,16$, $p > .10$, prin urmare ne vom putea încrede în prima valoare a lui t.

Ce înseamnă un $t(61) = 2,26$, $p < .05$? Rezultatul este semnificativ statistic, ceea ce duce la respingerea ipotezei nule care stipula că cele două populații nu diferă în privința nivelul de impulsivitate. Prin urmare, ipoteza cercetării este sprijinită, între cele două categorii de persoane existând diferențe ale nivelului de impulsivitate. Pentru a vedea direcția diferențelor este suficient să consultăm valorile mediilor prezentate în tabelul 3.5. Astfel, vom constata că persoanele care obişnuiesc să săreze mâncarea înainte de a o gusta sunt mai impulsive decât cele care sărează mâncarea după ce au gustat din ea.

Cele 61 de grade de libertate ale testului t independent din exemplul de mai sus pot fi calculate şi manual pe baza formulei ($df = n1 + n2 - 2$). Reamintim că cele două grupe au fost constituite din 36, respectiv 27 de persoane, fapt ce a condus la cele 61 de grade de libertate ale testului.

(6) Cât de intensă este diferența dintre grupe sau dintre categoriile de răspunsuri?

Mai rămâne să răspundem la o singură întrebare: cât de mare este această diferență între populații în ceea ce priveşte nivelul de impulsivitate? Pentru aceasta va trebui să calculăm mărimea efectului.

Spre deosebire de tehnicile χ^2 care sufereau la acest capitol, există o multitudine de indicatori ai mărimii efectului bazați pe stabilirea unor diferențe standardizate între medii. Dintre cei mai cunoscuți amintim de indicele d al lui Cohen sau g al lui Hedges. Mai multe informații despre aceşti coeficienți pot fi găsite în Cohen (1988) sau Rosenthal, Rosnow şi Rubin (2000a). Totuşi, pentru a nu suprasolicita cititorii cu explicarea unor noi indicatori, vom apela tot la coeficientul de determinare r^2 şi la corespondentul acestuia, coeficientul de corelație al mărimii efectului r, deşi nu ne aflăm în situația unor analize corelaționale. Formula de calcul a acestora este extrem de simplă:

$$r = \sqrt{\frac{t^2}{t^2 + df}} \qquad (3.3)$$

Pentru a afla valoarea lui r^2 vom utiliza aceeaşi formulă, dar fără radical. Prin urmare mărimea efectului se poate calcula simplu, pe baza valorii lui t şi a gradelor de libertate ale testului respectiv. Înlocuind în formulă ($t = 2.26$, iar $df = 61$), vom obține un r de .19, respectiv un r^2 de .035 (3,5%). Această valoare indică o diferență de magnitudine scăzută între cele două grupe cu privire la nivelul de impulsivitate.

Pentru a veni în sprijinul cititorilor familiarizați cu indicii d al lui Cohen, respectiv g al lui Hedges, precum şi pentru a facilita realizarea de meta-analize unitare, vom prezenta formulele de transformare a lui r în d, respectiv în g şi viceversa (Becker, 2003; Rosenthal, Rosnow şi Rubin, 2000a; Rosenthal, Rosnow şi Rubin, 2000b):

$$r = \frac{g}{\sqrt{g^2 + 4\left(\frac{df}{N}\right)}} \quad (3.4) \qquad g = \frac{\frac{r}{\sqrt{1-r^2}}}{\sqrt{\frac{df(n1+n2)}{n1n2}}} \quad (3.5) \qquad r = \frac{d}{\sqrt{d^2+4}} \quad (3.6)$$

$$d = \frac{2r}{\sqrt{1-r^2}} \quad (3.7) \qquad d = g\sqrt{\frac{N}{df}} \quad (3.8) \qquad g = d\sqrt{\frac{df}{N}} \quad (3.9)$$

unde r, g sau d reprezintă indicatorii mărimii efectului, N semnifică numărul total de subiecți testați, n1 şi n2 numărul de subiecți pentru fiecare din cele două grupe comparate, iar df reprezintă gradele de libertate. De remarcat că formulele se păstrează şi în cazul testelor ANOVA, cu precizarea că va fi vorba de gradele de libertate ale dispersiei neexplicate (df intragrup).

Iată şi demersul statistic în cazul problemei B:

Tabelul 3.6. Mediile şi abaterile standard ale mediilor şcolare anuale pentru cele două grupe

	Anul I de studii		Anul II de studii	
	M	AS	M	AS
Media generală	8,14	1,00	8,22	0,94

Este această diferență dintre medii, de doar opt sutimi, suficient de mare pentru a sprijini ideea studentului că anul II este mai uşor decât anul I? Pentru a răspunde la întrebare vom aplica testul t pentru două eşantioane perechi.

		Diferenta dintre medii	t	df	Sig. (2-tailed)
Pair 1	Media generala in anul I - Media generala in anul II	-,0814	-1,803	27	,082

Figura 3.9. *Informațiile esențiale din output-ul oferit de SPSS în cazul testului t pentru eşantioane perechi*
SPSS este o marcă înregistrată.

Rezultatul respectiv poate fi transcris în text sub forma t(27) = 1.80, p < .05, test unilateral. Se pot observa unele diferențe între datele prezentate în tabel şi forma uzuală de scriere a lui t. Astfel, în figura 3.9. t are semnul minus (-1.80). Această situație se explică prin faptul că în formula de calcul al lui t, diferența dintre cele două medii se găseşte la numărător. În cazul respectiv, restul este negativ,

deoarece media generală din anul I era mai mică decât media generală a anului II. Dacă am fi inversat termenii operației am fi obținut un rezultat pozitiv. În consecință, spre deosebire de corelații, în cazul testelor t şi z nu contează semnul acestora, ci doar rezultatul ca valoare absolută, în cazul de față 1,80. O altă deosebire față de tabel priveşte tipul de ipoteză testată. Din păcate, SPSS-ul oferă doar valoarea pragului de semnificație pentru testele bidirecționale. Dacă ipoteza cercetării este unidirecțională, precum în cazul problemei B, va trebui să calculăm manual pragul de semnificație pentru testul unilateral. Valoarea acestuia este la jumătate din cea a pragului unui test bidirecțional. În cazul de față, vom obține pentru testul unilateral un p de .041 (.082 / 2). Cu această ocazie putem constata importanța formulării ipotezei de cercetare. În cazul unui test bilateral („există diferențe semnificative între mediile şcolare din cei doi ani de studiu"), testul t nu ar fi fost semnificativ statistic, deoarece pragul obținut de .08 era mai mare decât pragul maxim admis de .05.

Prin urmare, datele nu ar fi sprijinit ideea unei diferențe în dificultate între cei doi ani universitari. În schimb, în cazul unui test unidirecțional („media generală din anul I va fi mai mică decât media generală din anul II"), testul t a devenit semnificativ statistic, deoarece un p de .04 este mai mic decât limita maximă de .05. Pe baza acestui rezultat şi analizând mediile din cele două testări, vom putea afirma că anul I de studiu este mai dificil decât anul al II-lea. Ținând cont că ipoteza problemei B este unidirecțională, a doua interpretare rămâne în picioare. Astfel, datele susțin intuiția studentului, rezultatele şcolare din primul an de studiu fiind mai slabe decât cele obținute în anul al II-lea.

Gradele de libertate în cazul testului t dependent se obțin pe baza formulei N-1, unde N reprezintă numărul de perechi de date. În cazul de față, au fost 28 de perechi (subiecți într-o grupă), prin urmare sunt 27 de grade de libertate.

Rămâne valabilă întrebarea: cât de mare este diferența dintre cele două medii? Aplicând aceleaşi formule prezentate în cazul testului independent vom obține un r^2 de 6,25%, corespunzător unui r de .25. Valoarea indică un efect de intensitate medie, existând diferențe destul de consistente între rezultatele şcolare obținute în cei doi ani. Anul I este mai dificil decât anul al II-lea. Pentru comparație, un efect scăzut corespunde expresiei „puțin mai dificil", iar un efect puternic, sintagmei „mult mai dificil".

Clarificări ale modului de constituire a celor două grupe comparate prin testul t sau z

Adesea se face confuzia că testele t sau z măsoară influența VI asupra VD. Acest lucru nu este întotdeauna adevărat. Pentru a întâlni acest aspect este nevoie, cel puțin, ca treptele VI să fie manipulate de către cercetător. Nu întotdeauna se întâmplă aşa. De multe ori criteriile care stau la baza constituirii celor două grupe nu pot fi manipulate de experimentator, ca în situația variabilelor etichetă. Exemple tipice ale acestora sunt sexul persoanei (există diferențe între

bărbați şi femei), grupa de vârsta (există diferențe între tineri şi vârstnici) etc. Mai mult, uneori criteriul de grupare poate fi o consecință a aşa-zisei variabile dependente. Obiceiul de a săra mâncarea înainte sau după gustarea mâncării pare a fi mai degrabă un efect al impulsivității şi nu o cauză a acesteia. Aceste subtilități joacă un rol important în interpretarea rezultatelor, precum identificarea corectă a relațiilor de tip cauză-efect.

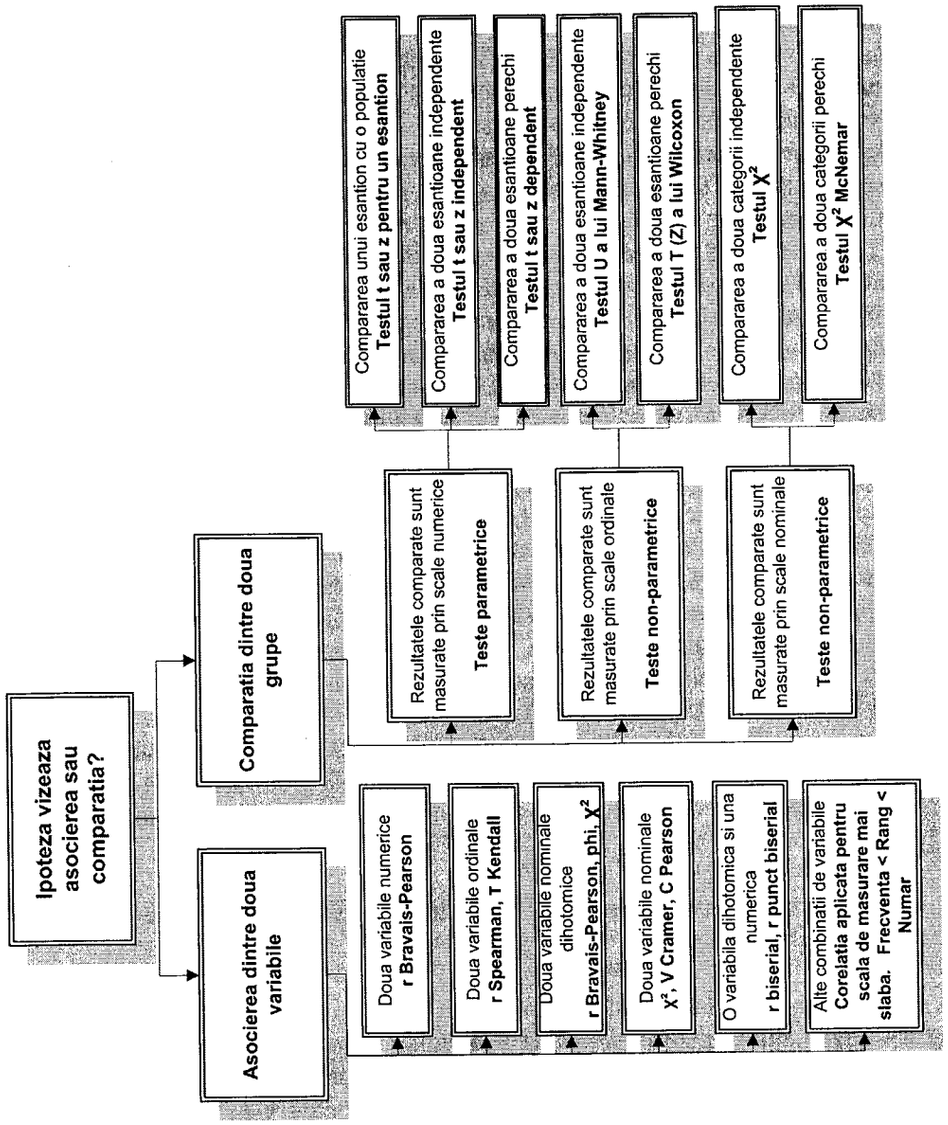

Ipoteza vizeaza asocierea sau comparatia?
Comparatia dintre doua grupe
Asocierea dintre doua variabile
Rezultatele comparate sunt masurate prin scale numerice
Teste parametrice
Rezultatele comparate sunt masurate prin scale ordinale
Teste non-parametrice
Rezultatele comparate sunt masurate prin scale nominale
Teste non-parametrice
Doua variabile numerice r Bravais-Pearson
Doua variabile ordinale r Spearman, T Kendall
Doua variabile nominale dihotomice r Bravais-Pearson, phi, X^2
Doua variabile nominale X^2, V Cramer, C Pearson
O variabila dihotomica si una numerica r biserial, r punct biserial
Alte combinatii de variabile Corelatia aplicata pentru scala de masurare mai slaba. Frecventa < Rang < Numar
Compararea unui esantion cu o populatie Testul t sau z pentru un esantion
Compararea a doua esantioane independente Testul t sau z independent
Compararea a doua esantioane perechi Testul t sau z dependent
Compararea a doua esantioane independente Testul U a lui Mann-Whitney
Compararea a doua esantioane perechi Testul T (Z) a lui Wilcoxon
Compararea a doua categorii independente Testul X^2
Compararea a doua categorii perechi Testul X^2 McNemar

Figura 3.10. *Ghid de alegere a testelor statistice inferențiale de bază*

Deşi testele t şi z sau tehnicile χ^2 sunt cele mai utilizate metode de compare a diferenței dintre două grupe, ele nu epuizează toate posibilitățile întâlnite. Figura 3.10 prezintă un ghid simplificat de alegere a testelor de comparație introductive, dar şi a celor de asociere. Ghidul îşi păstrează validitatea în cazul respectării condițiilor de aplicare a fiecărei tehnici statistice în parte (detalii despre toate aceste tehnici sau altele care le pot substitui pot fi găsite în Clocotici şi Stan (2000), Field (2000), Hinkle şi colab. (1994), Lungu (2001), Radu şi colab. (1993), Sava (2002) etc.).

3.3. Managementul datelor brute

În finalul acestui capitol vom discuta câteva măsuri care trebuie luate în cazul în care:

(1) există date lipsă;
(2) există date neobişnuite, extreme;
(3) datele brute trebuie transformate pentru a îndeplini condițiile de aplicare ale unor teste statistice parametrice.

3.3.1. Managementul datelor lipsă

Cu siguranță că în viața reală, oricât de mult ne-am dori noi ca subiecții testați să completeze toate rubricile, acest lucru are mici şanse să se întâmple din diverse motive: chestionare prea lungi, gradul de intimitate crescut al unor răspunsuri etc. În consecință, între rezultatele obținute vom regăsi şi spații libere, care vor constitui „cariile" bazei noastre de date. În astfel de cazuri avem la dispoziție mai multe strategii alternative de lucru (Tabachnick şi Fidell, 1996).

Prima alternativă, utilizată frecvent, este de a nu ține cont de datele lipsă şi de a elimina rezultatele subiectului respectiv. Tehnica poate fi aplicată în cazul în care sunt puține date lipsă, iar absența lor este aleatorie. Dacă sunt mai multe date lipsă, se pot realiza două grupe de subiecți: cei care au completat întreaga secvență dată şi cei care au date lipsă. Dacă există diferențe semnificative între cele două grupe, în cazul unor variabile importante pentru studiu, nu se recomandă ştergerea cazurilor cu date lipsă, deoarece se pierd informații importante.

În aceste condiții, cercetătorul va trebui să estimeze valoarea datelor lipsă şi să le introducă în baza de date. Pentru aceasta o măsură poate fi înlocuirea spațiilor goale cu una dintre valorile tendinței centrale (media, mediana sau modul), alegerea fiind făcută în funcție de tipul de date şi de forma distribuției acestora.

Totuşi, aşa cum vom vedea în secțiunea introductivă a capitolului destinat regresiei, media nu reprezintă cea mai bună modalitate de estimare a unei valori. Datele lipsă pot fi estimate mai bine printr-o ecuație de regresie. Variabila care include datele lipsă este utilizată drept criteriu, în timp ce variabilele care se

asociază foarte mult cu criteriul, vor fi luate drept predictori. Totuşi, estimarea datelor lipsă pe baza regresiei are şi ea anumite limite: valorile prezise pot depăşi amplitudinea unei variabile (ex. răspunsurile se regăsesc între 1 şi 5, iar valoarea estimată este 6); pentru aplicarea ei avem nevoie de variabile care să relaţioneze puternic cu variabila criteriu.

3.3.2. Managementul datelor neobişnuite

Am o veste bună pentru dumneavoastră: în medie, grupul de prieteni pe care-l aveţi câştigă, lunar, aproximativ 6.000 de USD ... însă cu o singură condiţie - să vă împrieteniţi cu un miliardar în dolari. Astfel, cele câteva milioane de lei pe care le câştigaţi dumneavoastră şi ceilalţi prieteni „obişnuiţi", adunaţi cu sute de mii de dolari câştigaţi de miliardar, vor da o medie care nu reprezintă pe nimeni – 6000 de USD. În cazul datelor bivariate, efectul distorsionant al datelor extreme a fost amintit în secţiunea destinată norului de puncte, când am prezentat efectul lor asupra valorii coeficientului de corelaţie.

Sursa valorilor neobişnuite poate consta în introducerea incorectă a unor date în calculator (de exemplu tastarea valorii 77, în locul lui 7) sau în includerea unui caz care nu face parte din populaţia de studiu, cum, din nefericire, este situaţia miliardarului în dolari. O altă cauză este existenţa unui număr mai mare de valori extreme decât cele aşteptate într-o distribuţie normală. Vă reamintiţi probabil, că 99% din datele distribuite normal au valori cuprinse între trei abateri standard distanţă, de o parte şi de alta a mediei. Astfel, dacă media unei variabile este 10, iar abaterea standard 1, ne aşteptăm ca 99% dintre persoane să obţină rezultate cuprinse între 7 şi 13. Dacă 7% din date se regăsesc în afara acestui interval, precum în cazul unei distribuţii platicurtice, vom avea mai multe valori extreme.

În funcţie de sursa datelor extreme, se impun şi măsurile corective. Astfel, dacă datele au fost introduse incorect, se poate face o verificare prin precizarea valorilor minime şi maxime întâlnite. Astfel, dacă datele variază de la 1 la 10, iar noi observăm că valoarea maximă este 66, cu siguranţă ea este o eroare de tehnoredactare, ce poate fi corectată. Dacă acel caz nu face parte din populaţia aflată în studiu, se elimină valoarea extremă prin renunţarea la cazul respectiv. În sfârşit, dacă ne aflăm în situaţia cu mai multe valori extreme decât sunt aşteptate într-o distribuţie normală, există două soluţii. Fie se realizează două categorii de rezultate, prin includerea sau excluderea valorilor neobişnuite, pentru a studia efectul acestora asupra indicatorilor statistici calculaţi, fie se va recurge la transformarea datelor brute. Ultima soluţie este utilă şi în cazul în care condiţiile de aplicare a unor teste statistice parametrice nu sunt îndeplinite.

3.3.3. Transformarea datelor brute

Vom preciza de la început că nu este vorba de o transformare a datelor brute în date standardizate de tipul cotelor T sau z, ci de o re-scalare a valorilor,

astfel încât condiţiile de aplicare ale unor teste statistice să fie îndeplinite. Acestea sunt discutate pe parcursul întregii lucrări, la momentul prezentării fiecărei tehnici statistice în parte. Deocamdată vom preciza doar că cele mai întâlnite condiţii sunt normalitatea distribuţiei, respectiv liniaritatea şi homoscedasticitatea relaţiei dintre două variabile.

Relaţiile neliniare pot fi tratate într-unul din următoarele moduri. Prima soluţie posibilă duce la gruparea valorilor variabilei X în trei categorii (mici, medii şi mari). O dată grupate, se pot aplica, fie tehnici de comparaţie a diferenţelor dintre cele trei grupe, fie tehnici de asociere între variabila Y şi fiecare variabilă de tip dummy, rezultată în urma grupării valorilor X. Detalii despre aceste tehnici vor fi oferite în cazul capitolelor referitoare la tehnicile ANOVA, respectiv în secţiunea despre variabile dummy din cadrul regresiei multiliniare. Cea de-a doua soluţie implică ridicarea la pătrat a valorilor X, obţinând astfel variabila X^2, care apoi va fi corelată cu variabila Y. Uneori, prin această măsură, relaţia dintre variabile devine liniară, însă acest lucru trebuie verificat prin inspectarea norului de puncte.

În cazul unor date numerice, homoscedasticitatea presupune ca împrăştierea punctelor să fie aproximativ aceeaşi pe toată distribuţia norului de puncte (finalul secţiunii 1.3.2), iar în cazul datelor nominale, implică ideea ca variabila dependentă să fie la fel de împrăştiată pe parcursul fiecărei categorii (omogenitatea dispersiei). Un exemplu de neîndeplinire a acestei condiţii în cazul datelor nominale se regăseşte în valorile de mai jos, deoarece datele sunt mult mai împrăştiate în cazul bărbaţilor decât în cazul femeilor:

Înălţime bărbaţi = {1,75; 1,80; 1,86; 1,67; 1,62; 1,70}

Înălţime femei = {1,66; 1,68; 1,70; 1,65; 1,63; 1,71}

În multe cazuri, heteroscedasticitatea apare ca urmare a faptului că una dintre variabile nu este distribuită normal. Tabachnick şi Fidell (1996) oferă o excelentă ilustrare pentru a susţine acest lucru, dând drept exemplu relaţia dintre venit şi vârstă. La vârste tinere, diferenţa dintre venituri nu este atât de mare, însă o dată cu înaintarea în vârstă, distanţele dintre oameni, în ceea ce priveşte venitul, cresc. Gradul de asociere dintre cele două variabile este atenuat de prezenţa heteroscedasticităţii. Venitul este probabil variabila nedistribuită normal, deoarece există mult mai multe valori mici şi medii, decât salarii foarte mari. Prin urmare, relaţia vârstă şi venit poate fi evaluată mai corect, dacă vom transforma iniţial valorile venitului, prin logaritmare. Vom lămuri imediat ce înseamnă acest lucru.

În sfârşit, normalitatea distribuţiei datelor este esenţială în obţinerea unor rezultate acurate pentru majoritatea tehnicile statistice, deoarece în realizarea lor s-a pornit de la premisa unei distribuţii normale. Când acest lucru nu se întâmplă, avem la dispoziţie două soluţii: fie renunţăm la aplicarea metodelor respective în favoarea unor tehnici non-parametrice, fie vom transforma datele pentru a normaliza, pe cât posibil, distribuţia datelor.

Pentru a înţelege ideea de transformare a datelor, vă punem următoarea întrebare: folosind acelaşi set de date, cum transformaţi o distribuţie non-normală

într-una normală? Altfel spus, cum puteți obține imaginea din dreapta figurii 3.11, pornind de la imaginea din stânga?

Figura 3.11. *Distribuție non-normală, platicurtică (stânga), respectiv o distribuție normală (dreapta)*

Este greu? Iată un alt indiciu! Să presupunem că imaginea din stânga este realizată pe baza următoarelor rezultate X = {2; 4; 7; 9; 12; 14; 16; 19; 22; 25}, astfel că valoarea minimă este 2, iar cea maximă 25. În schimb, imaginea din dreapta, are valori mult mai apropiate X ={1,41; 2; 2,64; 3; 3,46; 3,74; 4,35; 4,69; 5}, minimum fiind 1,41, iar maximum 5. Spiritul de observație şi abilitățile numerice v-au condus cu siguranță la răspunsul corect. Datele inițiale au fost transformate prin utilizarea radicalului. Astfel, o distribuție „împrăştiată" poate fi „adunată" prin transformarea valorilor observate, încât noile valori să aibă o amplitudine mai redusă.

Nu toate distribuțiile non-normale suferă de aceeaşi problemă, a unei distribuții platicurtice. Unele sunt puternic asimetrice pozitiv, altele sunt asimetrice negativ, iar în cazurile grave pot apărea atât devieri ale boltirii, cât şi ale oblicității. Pentru fiecare situație în parte există transformări specifice. Figura 3.12 oferă mai multe informații utile.

Transformările necesare sunt dependente de severitatea încălcării premisei de normalitate a distribuției. În cazul distribuțiilor asimetrice pozitiv putem scoate radical din valorile brute, putem logaritma sau putem obține inversul valorii. De exemplu, în cazul unei distribuții asimetrice pozitiv, care a deviat sever față de curba normală, datele pot fi transformate prin logaritmare. Aceleaşi operații pot fi efectuate şi în cazul asimetriilor negative, cu condiția ca transformarea să aibă loc utilizând datele reflectate. Reflectarea datelor se poate înțelege uşor prin următorul exemplu. Fie şirul de date: X = {3, 7, 9, 12, 14, 17}. Pentru a reflecta valorile, vom adăuga mai întâi cifra 1 la valoarea maximă din şir. Din valoarea rezultată, în cazul nostru 18, vom scădea fiecare valoare din şir, obținând X = {15, 11, 9, 6, 4, 1}.

Bineînțeles că acestea nu sunt singurele transformări posibile. Datele pot suferi diverse alte operații, precum ridicarea la pătrat, ridicarea la cub etc., însă soluțiile propuse mai sus dau cele mai bune rezultate în cazul unor deviații de la normalitate, asemenea celor prezentate în figura 3.12.

După transformare, în cazul normalizării aproximative a distribuției, datele pot fi utilizate, fără nici o problemă, în orice fel de tehnici statistice, care au drept premisă distribuția normală a datelor. Singurul aspect care se pierde prin transformare este unitatea de măsură a variabilei transformate. Astfel, dacă venitul

exprimat în euro are o distribuție asimetrică pozitiv care necesită logaritmare, valorile transformate reprezentând venitul vor avea ca unitate de măsură „euro logaritmați". În psihologie însă, multe rezultate de genul scorul 17 la un test de agresivitate, rezultatul 9 la o probă de sugestibilitate, nu au o unitate de măsură clară, aşa că nu vom pierde nici măcar atât prin transformarea datelor.

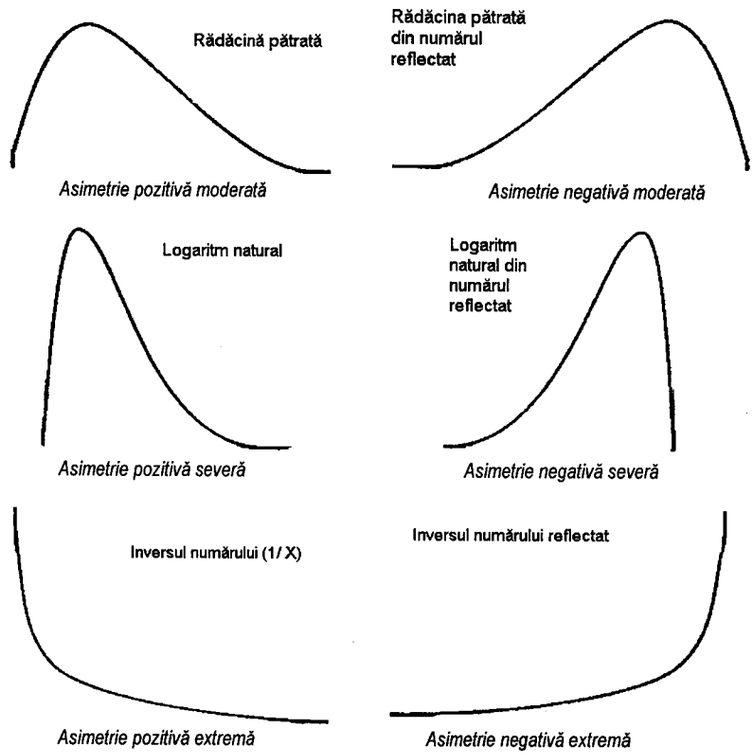


Figura 3.12. *Distribuții non-normale şi soluțiile corespunzătoare*

BIBLIOGRAFIE

APA (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

Becker, L. (2003). *Effect size.* Găsit la: http://www.uccs.edu/~lbecker/psy590/es.htm.

Brewer, M.B. (2000). Research design and issues of validity. În H.T. Reis şi C.M. Judd (eds.). *Handbook of Research Methods in Social and Personality Psychology* (p. 3-16). Cambridge: Cambridge University Press.

Cialdini, R.B., Borden, R.J., Thorne, A., Walker, M.R., Freeman, S., Sloan, L.R. (1999). Basking in reflected glory: three (football) field studies. În R.F. Baumeister (ed.). *The Self in Social Psychology* (p.436-445). Ann Arbor: Taylor & Francis.

Clocotici, V., Stan, A. (2000). *Statistică aplicată în psihologie.* Iaşi: Editura Polirom.

Cohen, J. (1988). *Statistical power analysis for the behavioral* sciences (2nd ed.). Hillsdale, USA: Erlbaum.

Faul, F., & Erdfelder, E. (1992). *GPOWER: A priori, post hoc, and compromise power analyses for MS-DOS [Computer program].* Bonn, Germany: Bonn University, Dept. of Psychology.

Field, A. (2000). *Discovering statistic using SPSS for Windows.* London: Sage Publications.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1994). *Applied statistics for the behavioral sciences* (3rd ed.). Boston: Houghton Mifflin Company.

Kerlinger, F.N, & Lee, H.B. (2000). *Foundations of behavioral research* (4th ed.). Orlando, USA: Harcourt College Publishers.

Leary, M.R., Tchividjian, L.R., & Kraxberger, B.E. (1999). Self-presentation can be hazardous to your health: impression management and health risk. În R.F. Baumeister (ed.). *The Self in Social Psychology* (p.182-194). Ann Arbor: Taylor & Francis.

Lungu, O. (2001). *Ghid introductiv pentru SPSS 10.0.* Seria Psihologie Experimentală şi Aplicată.

Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design, and analysis: an integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Radu, I., Miclea, M., Albu, M., Nemeş, S., Moldovan, O., Szamoskozi, Ş. (1993). *Metodologie psihologică şi analiza datelor.* Cluj-Napoca: Editura Sincron.

Rosenthal, R., Rosnow, R.L., & Rubin, D.B. (2000a). *Contrasts and effects sizes in behavioral research. A correlational approach.* Cambridge: Cambridge University Press.

Rosenthal, R., Rosnow, R.L., & Rubin, D.B. (2000b). Contrasts and correlations in effect-size estimation. *Psychological Science*, 11(6), 446-453.

Rotariu, T., Bădescu, G., Culic, I., Mezei, E., Mureşan, C. (1999). *Metode statistice aplicate în ştiinţele sociale.* Iaşi: Editura Polirom.

Sava, F. (2002). *Pagina de statistică socială.* Găsit la http://statisticasociala.tripod.com.

Tabachnick, B.G., Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins College Publishers.

Yaffee, R.A. (1996). A basic guide to statistical research and discovery. În F.T.L. Leong şi J.T. Austin (eds.). *The psychology research handook: a guide for graduate students and research assistants* (p.193-207). Thousand Oaks, USA: Sage Publications.

CAPITOLUL 4

TEHNICILE ANOVA

Când dorim să comparăm mai mult de două grupe de subiecți nu vom mai folosi testul t, ci vom apela la tehnicile ANOVA. ANOVA este un acronim de la denumirea în engleză „Analysis of Variance". În limba română tehnica este cunoscută şi sub denumirea de **analiză dispersională** sau **analiză de varianță.** Înainte de a aprofunda descrierea şi explicarea tehnicii ANOVA, voi prezenta motivele pentru care preferăm utilizarea ANOVA în locul testului t.

(1) Prima rațiune a preferinței pentru ANOVA ține de *design-ul cercetării.* Testul t se limitează la compararea diferențelor dintre două grupe, adică la situații în care există doar două trepte (modalități) ale variabilei independente. Orice manual de metodologie recomandă însă utilizarea unui număr cât mai mare de trepte ale variabilei independente pentru a observa mai bine efectul pe care factorii îl au asupra variabilei dependente. Pentru cei care simt nevoia unei clarificări suplimentare, voi apela la un exemplu bazat pe legea Yerkes-Dodson, care face referire la relația dintre motivație şi performanță. Se cunoaşte că există un optim motivațional care va duce la performanțe crescute. În schimb, dacă subiecții sunt submotivați sau supramotivați vor obține performanțe scăzute. Pornind de la aceste considerente, să ne imaginăm situația în care am aplica testul t pentru două grupe de subiecți, cei care au o motivație scăzută (submotivați) şi cei care au o motivație foarte ridicată (supramotivați). În aceste condiții nu vom observa diferențe între cele două grupe în ceea ce priveşte nivelul lor de performanță şi vom trage concluzia că motivația nu afectează performanța. În schimb, dacă vom introduce cel puțin o grupă suplimentară, formată din subiecți optim motivați, vom observa diferențe semnificative între cele trei grupe şi vom trage concluzia că motivația influențează nivelul de performanță. Astfel, cu cât numărul de grupe de comparație este mai mare, cu atât se diversifică treptele manipulate ale variabilei independente şi vom observa mai bine influența variabilei independente asupra celei dependente.

(2) Cel de-al doilea motiv pentru care testul t nu este adecvat în situații de comparare a mai mult de două grupe este de ordin *statistic* şi se numeşte eroarea setului de comparații.

Pentru a înțelege empiric acest concept, iată cum ar sta lucrurile pentru situația de mai sus, referitoare la influența motivației asupra performanței. Utilizând testul t de comparații între două grupe vom obține trei indicatori t, corespunzători comparării grupei 1 cu grupa 2, grupei 1 cu grupa 3 şi grupei 2 cu grupa 3. Utilizând ANOVA, vom fi mult mai eficienți, obținând un singur indicator F. Acesta nu este însă singurul neajuns. Nu este doar o problemă de a munci de trei ori mai

mult, ci de prezența unor distorsiuni declanşate prin calcularea mai multor teste t, un caz particular al erorii setului de comparații.

Aceasta este determinată de formula următoare (Hinkle, Wiersma şi Jurs, 1994):

$$1-(1-\alpha)^c \qquad (4.1)$$

unde α este pragul de semnificație stabilit a priori pentru fiecare test t efectuat (frecvent p=.05), iar c este dat de numărul de teste t ce trebuie efectuate.

Dacă există doar două grupe de subiecți, se aplică un singur test t pentru a testa diferențele dintre mediile celor două populații de subiecți (a). Dacă există trei grupe de subiecți, precum în situația prezentată mai sus, vom avea de aplicat trei teste t (b). În cazul testării diferențelor dintre patru grupe am fi aplicat şase teste t (1 cu 2; 1 cu 3; 1 cu 4; 2 cu 3; 2 cu 4; şi 3 cu 4) (c). Aplicând formula de mai sus, vom observa, însă, că doar în primul caz (un singur test t pentru un design cu două grupe), probabilitatea de a comite o eroare de tip I rămâne .05.

$$1-(1-0.05)^1 = 0.05(a)$$

$$1-(1-0.05)^3 = 0.15(b)$$

$$1-(1-0.05)^6 = 0.27(c)$$

În situațiile în care comparăm mai mult de două grupe, eroarea de tip I depăşeşte limita maximă acceptată de .05, apărând eroarea setului de comparații. Pentru situația b, de comparare a trei grupe, ea se citeşte în felul următor: există o probabilitate de 15% de a comite eroarea de tip I într-una din cele trei situații de comparare. Astfel, noi vom respinge ipoteza nulă când ea nu ar fi trebuit respinsă. Similar, pentru situația c, de comparare a patru grupe, vom avea o probabilitate de 27% să obținem o eroare de tip I într-una din cele şase comparații făcute.

Asemenea situații sunt inacceptabile, deoarece pragul maxim al probabilității de apariție a erorii de tip I este de 5% (.05). Astfel, nu vom putea utiliza testul t în situații de comparare a mai mult de două grupe.

Chiar şi în aceste condiții, unii autori nu renunță la ideea utilizării testelor t şi efectuează o ajustare, prin metoda Bonferroni, astfel încât să elimine eroarea setului de comparații şi să controleze şansele de a comite eroarea de tip I. Ajustarea este simplă şi constă în împărțirea pragului de semnificație de .05 la numărul de teste t care urmează a fi aplicat. În cazul unui design cu patru grupe, fiind necesare şase teste t, vom obține un nou prag de semnificație pentru fiecare din cele şase teste t de .008. O asemenea valoare, deşi controlează şansa de a comite eroarea de tip I, este foarte conservatoare, crescând şansele de a comite eroarea de tip II, şi anume, de a nu respinge ipoteza nulă, când ar fi trebuit respinsă, deoarece există diferențe semnificative între populațiile din care sunt extrase grupele.

Prin urmare, în afara obişnuinței sau uşurinței de a aplica testul t, nu există nici un motiv temeinic pentru păstrarea testelor t în situații în care comparăm mai mult de două grupe. Sintetizând, ele nu sunt deloc economice (ex. design-ul cu cinci grupe necesită zece teste t, în timp ce ANOVA necesită un singur indicator F); cresc şansele de a comite eroarea de tip I, iar atunci când aceasta este controlată, sporesc şansele de comitere a erorii de tip II.

Tipologia tehnicilor ANOVA

Cele mai întâlnite design-uri de cercetare experimentală utilizează una dintre următoarele patru forme de analiză dispersională:

a) ANOVA simplă (unifactorială; o cale);
b) ANOVA cu măsurători repetate;
c) ANOVA factorială (ex. bifactorială; două căi);
d) ANOVA mixtă.

Iată o scurtă descriere a acestor patru tipuri:

(a) ANOVA simplă este modelul cel mai simplu dintre tehnicile ANOVA, fiind un corespondent al testului t pentru două eşantioane independente.

> *Exemplu:* Trei grupe de subiecți sunt obținute în mod aleatoriu pentru a testa efectul pe care-l are „Viagra" asupra potenței sexuale. O grupă de subiecți va primi „placebo", pentru a controla efectul auto-sugestiei; o altă grupă va primi o doză mică de „Viagra", în timp ce a treia grupă primeşte o doză ridicată de „Viagra". Variabila dependentă a fost nivelul potenței sexuale (Field, 2000).

(b) ANOVA cu măsurători repetate este un model corespondent al testului t pentru două eşantioane perechi.

> *Exemplu:* Zece subiecți se oferă voluntar pentru a testa efectul pe care-l are „Viagra" asupra potenței sexuale. Mai întâi subiecții vor primi „placebo", pentru a controla efectul auto-sugestiei, după care li se va măsura potența sexuală. Apoi, aceiaşi subiecți vor primi „Viagra" în doze mici, după care va fi măsurată din nou potența sexuală. În sfârşit, aceiaşi subiecți vor primi o doză mare de „Viagra", după care vor fi măsurate efectele apărute la nivelul potenței sexuale.

(c) ANOVA factorială este un model mai complex, fără un echivalent direct în cadrul testelor t. ANOVA factorială se distinge prin faptul că vor fi testate efectele mai multor variabile independente (factori) asupra unei singure variabile dependente, precum în cazul unui design factorial.

> *Exemplu:* Vom păstra parțial situația din exemplul ANOVA simplă, referitor la manipularea variabilei independente „Viagra" cu trei trepte (modalități): placebo; doză redusă; doză mare. Participanții la experiment vor fi împărțiți aleatoriu într-una din cele trei situații experimentale. Dacă credem că nivelul de stres al participanților influențează potența sexuală, vom putea introduce o altă

variabilă independentă cu două trepte (stres ridicat; stres scăzut) şi vom putea împărți participanții în două categorii după acest criteriu. În consecință, vor rezulta şase grupe experimentale, într-un design factorial de tip 3 X 2. Iată cele şase situații experimentale posibile:

Placebo x stres scăzut	Viagra, doză mică x stres scăzut	Viagra, doză mare x stres scăzut
Placebo x stres ridicat	Viagra, doză mică x stres ridicat	Viagra, doză mare x stres ridicat

Fără a mai exemplifica, ANOVA factorială poate fi utilizată în cazul măsurătorilor repetate dacă sunt manipulați cel puțin doi factori.

(d) ANOVA mixtă presupune o combinație între ANOVA simplă şi ANOVA cu măsurători repetate.

Exemplu: Similar exemplului oferit în cazul tehnicii ANOVA simplă, manipularea variabilei independente „Viagra" cu trei trepte (placebo, doză redusă, doză mare) se face distribuind aleatoriu participanții într-una din cele trei situații experimentale. În schimb, vom manipula diferit variabila „stres". Vom utiliza aceiaşi subiecți în două situații diferite, reprezentând cazuri de stres ridicat sau stres scăzut. Stresul ridicat a fost reprezentat de perioada de lucru intens, când persoanele au fost foarte ocupate şi stresate de termenele de finalizare a diferitelor proiecte, iar stresul scăzut a fost manipulat pe aceleaşi persoane, în situații de odihnă, reprezentate de perioada de concediu. Astfel, deşi obținem tot şase celule experimentale ca în cazul ANOVA factorială, nu este vorba şi de şase grupe de subiecți diferiți. În această situație vor fi doar trei grupe de subiecți diferiți (placebo, doză mică, doză mare), iar fiecare din aceştia va fi testat cu două ocazii diferite (stres ridicat – suprasolicitare plus presiunea termenelor limită; respectiv stres scăzut – concediu de odihnă).

4.1. ANOVA simplă

Tehnica permite compararea simultană a trei sau mai multe trepte (modalități) ale unei singure variabile independente, menținând nivelul alfa la valoarea dorită, de maximum .05. O modalitate clasică de analiză matematică a acestei tehnici este analiza dispersiei variabilei dependente. Dispersia totală a acesteia poate fi împărțită în două componente: dispersia dintre grupe (intergrup) şi dispersia din interiorul grupelor (intragrup). Clocotici şi Stan (2000) denumesc aceste surse variația dintre grupuri, respectiv variația în grupuri. În cadrul ANOVA există două categorii de medii: cea a răspunsurilor pentru fiecare grupă în parte şi marea medie, rezultată din luarea în considerare a tuturor răspunsurilor, indiferent de grupa din care fac subiecții parte. Spre exemplu, subiectul 1 din grupa 1 are o anumită abatere față de media răspunsurilor din grupa respectivă şi o altă abatere

față de marea medie, rezultată din toate răspunsurile exprimate în cercetare. Pentru a putea distinge mai bine între grupuri, ar trebui ca prima componentă a variației să fie mai mică decât cea de-a doua. Prima abatere intră în sfera dispersiei intragrup şi se datorează variațiilor individuale, determinate de gruparea aleatoare a subiecților. Cea de-a doua abatere intră în sfera dispersiei intergrup, care reflectă atât gruparea aleatoare a subiecților, cât mai ales fluctuațiile datorate influenței variabilei independente asupra celei dependente. Simplificând lucrurile, dacă dispersia intergrup este mult mai mare decât cea intragrup, există diferențe semnificative între grupele comparate, deci există o influență a variabilei independente asupra celei dependente.

Totuşi, trebuie precizat faptul că ANOVA simplă nu constituie un panaceu universal pentru orice situație. Ea are două **limite** importante:

1) poate fi aplicată, asemenea celorlalte tehnici ANOVA, doar în cazul în care treptele variabilei independente sunt de tip nominal (ex. bărbați vs. femei) sau de tip ordinal (placebo, doză mică, doză mare). În plus, variabila dependentă este măsurată printr-o scală de interval sau proporții, care îndeplineşte câteva condiții ce vor fi prezentate ulterior;
2) rezultatul final, exprimat printr-o valoare F, exprimă un scor general, fără a preciza specificul evoluției comparației dintre grupuri.

Pentru a înțelege mai bine cel de-al doilea aspect vom face o scurtă incursiune în testarea ipotezei statistice pentru testele t şi tehnicile ANOVA.

În cazul testelor t, ipoteza nulă este Ho: $\mu 1 = \mu 2$. Respingerea ipotezei nule însemna automat că cele două grupe diferă semnificativ în ceea ce priveşte mediile lor ($\mu 1 \neq \mu 2$).

În cazul ANOVA simplă pentru trei grupe, ipoteza nulă este Ho: $\mu 1 = \mu 2 = \mu 3$. Respingerea ipotezei nule ne arată că cele trei medii nu sunt egale, fără a ne spune însă cum diferă ele în mod specific. Astfel, rămân câteva alternative valabile, de genul $\mu 1 \neq \mu 2 \neq \mu 3$ sau $\mu 1 = \mu 2$, dar $\mu 2 \neq \mu 3$ etc. La aceste aspecte nu putem răspunde prin testul F, deoarece acesta rămâne la un nivel general de testare a ipotezei statistice, enunțat în ipoteza nulă.

Pornind de la acest inconvenient, există două strategii de rezolvare a problemei:

- utilizarea comparațiilor a priori, de genul contrastelor polinomiale, a celor ortogonale sau a celor non-ortogonale, fără a ne baza prea mult pe calcularea indicatorului general F;
- utilizarea unor comparații post hoc, între anumite grupe specifice, efectuate în cazul în care F a fost semnificativ statistic, indicând existența unor diferențe între grupele cercetării. Probele post hoc au calitatea de a păstra, în limitele acceptabile ale pragului de .05, riscul de a comite o eroare de tip I şi, totodată, de a identifica în mod specific care dintre grupe sunt diferite şi au condus la respingerea ipotezei nule.

Dintre cele două modalități de acțiune prima este mai elegantă şi mai capabilă de a detecta diferențele semnificative dintre grupe, însă cea de a doua

este mai populară, majoritatea cercetărilor apelând, din tradiție, la tehnici post hoc. Vom prezenta, pe rând, ambele modalități de lucru, însă vom începe cu cea din urmă datorită simplității sale.

Iată exemplul utilizat pentru a explica ambele modalități de lucru:

Un behaviorist doreşte să vadă care este eficiența diferitelor tehnici terapeutice comportamentale în tratarea fricii de vampiri. Au fost cuprinse 36 de persoane voluntare în studiu, care declarau anterior credința lor în existența vampirilor şi care manifestau o fobie crescută față de aceştia. Participanții au fost împărțiți, în mod aleatoriu, în trei grupe. Grupa repartizată tehnicii „in vivo" a petrecut, timp de o săptămână, nopțile într-un motel aflat în vecinătatea unui cimitir părăsit. Grupa repartizată tehnicii de „desensibilizare sistematică" a învățat mai întâi o tehnică de relaxare, după care a vizionat, pe parcursul unei săptămâni, câteva filme cu vampiri, pe fondul unei muzici de relaxare şi pe fondul exersării tehnicilor învățate cu scopul de a reduce frica de vampiri. Cel de-al treilea grup a fost păstrat cu scop de control. La finalul săptămânii toți participanții au fost testați printr-o scală specifică de evaluare a fricii de vampiri. Rezultatele mici indică o fobie scăzută, în timp ce rezultatele mari indică un nivel crescut de fobie. Există diferențe semnificative între populațiile reprezentate prin cele trei grupuri cu privire la acest aspect?

Baza de date aferentă acestui exemplu este anova simpla.sav. În SPSS, ANOVA simplă poate fi obținută urmând una dintre cele două căi, fie „Analyze" - „Compare Means" - „One-way ANOVA", fie „Analyze" - „General Linear Model" - „Univariate".

4.1.1. Procedurile post hoc

În cazul utilizării procedurilor post hoc, rezolvarea problemei presupune urmarea a doi paşi:

a) calcularea lui F pentru a vedea dacă există diferențe semnificative între grupe, cu alte cuvinte, pentru a vedea dacă terapiile comportamentale sunt eficiente în reducerea fobiilor;
b) alegerea testelor post hoc adecvate pentru a vedea cum diferă grupele între ele, pas necesar a fi făcut doar în condițiile respingerii ipotezei nule din primul pas.

ANOVA

frica de vampiri

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	160,167	2	80,083	50,986	,000
Within Groups	51,833	33	1,571		
Total	212,000	35			

Figura 4.1. *Output SPSS cu modul clasic de prezentare a testului F în tabelul sumar ANOVA*

SPSS este o marcă înregistrată.

Iată o scurtă explicitare a tabelului sumar ANOVA din figura 4.1:

- sursa totală a dispersiei (prima coloană) este constituită din dispersia intergrup şi dispersia intragrup;
- SS sau suma pătratelor (a doua coloană) este o sumă a abaterilor pentru ambele situații;
- gradele de libertate sunt prezentate în a treia coloană. Pentru dispersia intergrup ele se calculează pe baza numărului de grupe minus unu (3-1), în timp ce pentru dispersia intragrup ele sunt egale cu numărul total de subiecți minus numărul de grupe (36-3);
- media pătratică este rezultatul raportului dintre suma pătratelor şi gradele de libertate pentru fiecare tip de dispersie în parte. Mai mult, prin raportul celor două medii pătratice se obține rezultatul final, F;
- F se interpretează într-un mod similar altor teste statistice. Dacă valoarea sa este semnificativă statistic ($p < .05$), atunci vom respinge ipoteza nulă conform căreia nu există diferențe între populațiile din care au fost extrase grupele implicate în design-ul experimental;
- rezultatul lui F se scrie în felul următor: $F(2, 33) = 50,98$, $p < .01$.

În asemenea situații am putea afirma că, diferențele dintre cele trei grupe nu se datorează hazardului. Probabil, intervențiile terapeutice de sorginte comportamentală au efect asupra reducerii fobiei de vampiri la subiecții incluşi în studiu. Pentru a observa cum diferă cele trei grupe vom apela la metodele post hoc.

Dintre multitudinea procedurilor dezvoltate în acest scop, vom folosi două criterii importante în alegerea tehnicilor adecvate:

a) există un număr de subiecți egali în grupele cercetării?
b) există o anumită omogenitate a dispersiei între aceste grupe?

Tabelul 4.1. Modul de selecție a testelor post hoc în funcție de cele două criterii

		Număr de subiecți		
		Egali (ex. 10, 10, 10)	Relativ inegali (ex. 11, 10, 9)	Profund inegali (ex. 15, 10, 7)
Omogenitatea dispersiei	Dispersii egale	*Tukey*	*Gabriel* *Tukey b*	*Hochberg GT2*
	Dispersii inegale	*Dunnett T3*	*Games-Howell*	*Games-Howell*

Dacă numărul de subiecți este un criteriu evident încă din datele inițiale ale unei cercetări, omogenitatea dispersiei se testează prin apelul la testul Levene. În situația în care rezultatul este nesemnificativ statistic, dispersiile sunt considerate a fi egale. Dacă rezultatul este semnificativ statistic, dispersiile sunt considerate a fi eterogene, diferite. În această situație, vom încerca mai întâi să eliminăm problema prin transformarea datelor variabilei dependente. Dacă o asemenea soluție nu va îndrepta situația, vom apela la metodele de comparare post hoc precizate în cazul

dispersiilor inegale, după ce, în prealabil, vom ajusta citirea lui F. Modul de lucru în acest caz va fi specificat ulterior, în descrierea condițiilor necesare aplicării testului F.

Multiple Comparisons

Dependent Variable: frica de vampiri

	(I) situatie experimentala	(J) situatie experimentala	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	in vivo	desensibilizare sistematica	-1,08	,512	,102
		control	-4,92*	,512	,000
	desensibilizare sistematica	in vivo	1,08	,512	,102
		control	-3,83*	,512	,000
	control	in vivo	4,92*	,512	,000
		desensibilizare sistematica	3,83*	,512	,000
Dunnett T3	in vivo	desensibilizare sistematica	-1,08	,514	,131
		control	-4,92*	,529	,000
	desensibilizare sistematica	in vivo	1,08	,514	,131
		control	-3,83*	,491	,000
	control	in vivo	4,92*	,529	,000
		desensibilizare sistematica	3,83*	,491	,000

*. The mean difference is significant at the .05 level.

Figura 4.2. *Rezultatele comparațiilor post hoc oferite de SPSS 11.0*
SPSS este o marcă înregistrată.

Pentru problema de față, testul Levene a fost nesemnificativ statistic, iar numărul de subiecți din fiecare grupă a fost egal, fiind câte 12 participanți în fiecare situație experimentală. În aceste condiții, testul Tukey este cea mai potrivită procedură de calcul a diferențelor specifice dintre cele trei grupe. Oricare altă procedură, dintre cele prezentate în tabelul de mai sus, ar fi neadecvate în această situație, deoarece ar fi prea liberale (nu ar respecta condiția ca probabilitatea de eroare de tip I să fie de maximum 5%) sau prea conservatoare (cresc şansele de a comite eroarea de tip II). Totuşi, datorită faptului că omogenitatea dispersiei nu poate fi cunoscută înaintea rulării programului statistic, pot fi selectate mai multe proceduri post hoc cu mențiunea că vom lua în considerare rezultatele obținute prin tehnica ce respectă cel mai bine condițiile întâlnite în cazul respectiv, cum ar fi testul Tukey pentru problema de față.

Urmărind valorile testului Tukey vom putea decela următoarele aspecte:

- nu există o diferență semnificativă în ceea ce priveşte reducerea fricii de vampiri între persoanele care au beneficiat de o terapie in vivo față de cele care au învățat desensibilizarea sistematică (Tukey = -1,08, p >.10);

- terapia comportamentală bazată pe tehnica in vivo este eficientă în reducerea fricii de vampiri, rezultatele subiecților din această grupă fiind semnificativ mai mici decât cele obținute de grupul de control, formată din persoane care nu au urmat nici un fel de terapie (Tukey = -4,92, p < .01);
- similar, terapia comportamentală bazată pe tehnica desensibilizării sistematice este eficientă în reducerea fricii de vampiri, rezultatele subiecților din această grupă fiind semnificativ mai mici decât cele obținute de grupul de control (Tukey = -3,83, p < .01);
- pentru cei comozi, SPSS-ul oferă un sprijin important în a identifica diferențele semnificative la un p < .05, acestea fiind marcate cu asterisc. Totuşi, merită remarcat şi reversul acestei situații. Valorile testului Tukey nu sunt specificate direct, însă ele pot fi calculate pe baza datelor din figura 4.2., făcând un raport între diferența dintre medii şi eroarea standard a mediilor. De exemplu, valoarea testului Tukey în prima situație nu este -1,08, ci 1,99 (1,08 / 0,512). Similar, valoarea lui t Tukey pentru comparația dintre subiecții supuşi terapiei in vivo şi cei din grupul de control a fost 9,60 (4,92 / 0,512), iar pentru comparația dintre subiecții supuşi desensibilizării sistematice şi grupul de control a fost 7,48 (3,83 / 0,512).
- ultimă observație priveşte semnul testelor post hoc. Asemenea testelor t, contează valoarea absolută obținută, fără a ține cont de semn. Semnul minus nu înseamnă altceva decât faptul că subiecții au obținut, în medie, rezultate mai mari la a doua grupă de comparație decât la prima.

Aspectele enunțate mai sus pot fi sintetizate astfel: în urma aplicării testului Tukey, ambele forme de terapie comportamentală s-au dovedit a reduce semnificativ frica de vampiri, prin comparare cu grupul de control. În schimb, nu putem afirma că una dintre cele două forma de terapie, in vivo sau desensibilizare sistematică, este mai eficientă în reducerea fricii de vampiri.

4.1.1.1. Mărimea efectului în cazul procedurilor post hoc

Prin testul de semnificație statistică am reuşit să arătăm o anumită influență a tipurilor de terapie asupra reducerii fobiei de vampiri. Rămâne în continuare să stabilim cât de mare este această reducere? Cât de eficiente sunt tehnicile terapeutice în comparație cu lipsa oricărei intervenții specifice grupului de control? Pentru a oferi un răspuns la aceste întrebări vom calcula mărimea efectului, singura modalitate prin care putem decide importanța practică a rezultatelor obținute. După cum am demonstrat atunci când ne-am referit la mărimea efectului, un rezultat semnificativ statistic la un p < .05 poate fi lipsit de importanța practică în cazul unui efect minuscul.

Modul de calculare a acestuia se poate face utilizând una din cele două formule (Rosenthal, Rosnow şi Rubin, 2000a; Rosenthal, Rosnow şi Rubin, 2000b):

a) în cazul în care VI are doar două trepte (modalități):

$$r = \sqrt{\frac{F}{F + df_{\text{int ragrup}}}} \qquad (4.2)$$

b) în cazul în care VI are mai mult de două trepte (modalități) şi s-a procedat la comparații post hoc sau la contraste:

$$r = \sqrt{\frac{F_{contrast}}{F * df_{\text{int ergrup}} + df_{\text{int ragrup}}}} \qquad (4.3)$$

F reprezintă valoarea testului omnibuz, prin care se stabileşte dacă ipoteza nulă va fi sau nu respinsă. F contrast este valoarea obținută în cadrul unui anumit contrast sau comparație post hoc, ştiindu-se faptul că $t^2 = F$.

În cazul nostru, fiind vorba de trei comparații post hoc, putem calcula trei mărimi ale efectelor. Totuşi, datorită faptului că nu au fost observate diferențe semnificative între cele două intervenții psihoterapeutice nu vom calcula mărimea efectului în acel caz, deoarece diferențele observate s-ar putea datora hazardului într-o mai mare măsură. Acest lucru nu este obligatoriu, mărimea efectului putând fi calculată şi în condițiile în care un rezultat nu a fost semnificativ statistic. În cazul unui număr relativ mic de subiecți şi în situația unei mărimi a efectului ridicate, un rezultat nesemnificativ statistic ar fi cauza unei puteri reduse de a decela diferențele semnificative şi nu a lipsei de influență sau asociere între cele două variabile în cauză.

În cazul nostru, obținând valorile F pe baza rezultatelor prezentate în figura 4.2. şi a faptului că $t^2 = F$, vom utiliza formula 4.3 pentru a obține mărimea efectului pentru terapia in vivo în comparație cu rezultatele grupului de control:

$$r = \sqrt{\frac{9{,}60^2}{50{,}98 * 2 + 33}} = 0{,}82$$

Într-un mod similar, mărimea efectului intervenției bazate pe desensibilizare sistematică în comparație cu grupul de control a fost r = 0,64. Dacă ar fi să ridicăm aceste valori la pătrat pentru a observa procentul din dispersia VD influențat de VI în cauză vom obține 67%, respectiv 0,40%. Ambele valori sunt foarte ridicate, indicând un tratament excelent al fobiei de vampiri prin cele două metode behavioriste. Evaluarea parametrilor s-a făcut pe baza tabelului 2.2. prezentat în secțiunea destinată mărimii efectului.

În final, doresc să amintesc că valorile r sau r^2 pot fi transformate în alți indicatori ai mărimii efectului pe baza formulelor 3.4-3.9 şi că SPSS-ul oferă o variantă de mărimea efectului denumită eta pătrat (η^2). Aceasta nu poate pătrunde la nivelul analitic, al comparațiilor dintre două grupe, fiind doar un estimator bazat

pe F omnibuz. Valoarea poate fi utilă doar în condițiile în care VI are doar două trepte, însă, chiar şi în acele condiții, poate atinge valori eronate.

4.1.2. Procedurile a priori

În cazul în care există o anumită direcție a ipotezei statistice, ar fi mai util să se apeleze la tehnicile a priori (contraste), care, în comparație cu testul general F, au o mai mare putere de decelare a influenței variabilei independente asupra celei dependente. Astfel, dacă există vreun efect, acesta este mai uşor decelat prin apelul la contraste decât prin metoda clasică în doi paşi, de calculare a lui F, urmată de alegerea unui test post hoc.

Contrastele sunt proceduri statistice care testează ipoteze specifice. Prin aceasta se deosebesc de testul general F, prin faptul că oferă mai multă claritate conceptuală. Mai mult, contrastele stau la baza calculării magnitudinii influenței variabilei independente asupra celei dependente (Rosenthal şi colab., 2000a).

Există mai multe modalități de lucru, implicit mai multe tipuri de contraste. Acestea pot fi grupate după două criterii majore: gradul lor de standardizare şi prezența sau absența ortogonalității. Astfel, avem contraste nestandardizate ortogonale sau non-ortogonale, respectiv contraste standardizate ortogonale sau non-ortogonale. Numele de ortogonal vine de la faptul că setul de ipoteze este independent, acestea neinfluențându-se reciproc. Astfel, eroarea setului de comparații rămâne constantă la nivelul alfa dorit de .05 sau .01. În plus, o categorie aparte de contraste standardizate sunt cele polinomiale, care realizează o analiză direcțională. Toate aceste aspecte vor fi discutate pe rând.

4.1.2.1. Contraste nestandardizate

Pentru a înțelege mecanismul de funcționare şi alegere a contrastelor vom arăta modul în care ar putea fi împărțită dispersia totală, bazându-ne pe exemplul anterior legat de influența terapiilor comportamentale asupra reducerii fricii de vampiri.

Dispersia totală = dispersia explicată (intergrup) + dispersia erorii (intragrup)

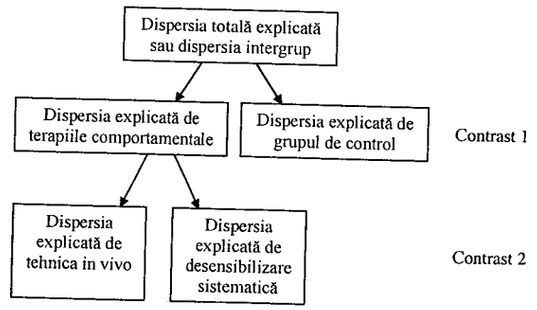

Dispersia totală explicată sau dispersia intergrup
Dispersia explicată de terapiile comportamentale
Dispersia explicată de grupul de control
Contrast 1
Dispersia explicată de tehnica in vivo
Dispersia explicată de desensibilizare sistematică
Contrast 2

Figura 4.3. *Exemplu de partiționare a dispersiei în cazul contrastelor ortogonale*

Avantajul contrastelor este de a putea diviza dispersia intergrup în unități componente de interes pentru cercetător, fiind astfel mult mai specifică decât cazul testului general F. Spre exemplu, în problema noastră ar fi interesant să stabilim dacă fobia de vampiri se reduce semnificativ prin apelul la oricare tehnică comportamentală în comparație cu grupul de control (contrast 1). Similar, am putea fi interesați să vedem dacă există diferențe semnificative între cele două modalități de terapie comportamentală în ceea ce privește eficiența lor în reducerea fobiei de vampiri (contrast 2). Aceste aspecte pot decurge din dispersia totală la fel ca în reprezentarea grafică de mai sus.

În realizarea contrastelor nestandardizate ortogonale sunt totuși câteva reguli ce trebuie respectate:

a) numărul de contraste realizate este întotdeauna k-1, unde k reprezintă numărul total de grupe ce pot fi comparate. În exemplul nostru, existând trei grupe denumite generic „de control", „desensibilizare sistematică" şi „in vivo", pot fi realizate maximum două contraste;
b) dacă un grup a fost îndepărtat de celelalte grupe pentru a fi comparat cu acestea (ex. cazul grupului de control în primul contrast), el nu mai poate fi utilizat în contrastele ulterioare;
c) grupele vor fi ponderate, astfel încât să existe o egalitate de coeficienți în fiecare contrast efectuat.

Această ponderare se face ținând seama de numărul de subiecți dintr-o grupă şi de numărul de grupe dintr-un „chunk". În situații de grupe cu număr egal de subiecți, ca în exemplul nostru, coeficienții atribuiți pentru cele două contraste de mai sus pot fi:

	„in vivo"	desensibilizare sistematică	control
Contrast 1	1	1	-2
Contrast 2	1	-1	0

La fel de bine, aceşti coeficienți pot fi:

	„in vivo"	desensibilizare sistematică	control
Contrast 1	1/2	1/2	-1
Contrast 2	1	-1	0

Pentru a determina dacă toți coeficienții sunt atribuiți corect, trebuie ca oricare sumă realizată pe orizontală să fie zero (ex. 1 + 1 -2 = 0; 1 – 1 + 0 = 0 sau ½ + ½ -1 = 0). Pentru a determina dacă contrastele îndeplinesc criteriul contrastelor ortogonale ar trebui ca suma produselor a două contraste să fie tot zero. Spre exemplu, în primul caz avem: 1*1 + 1*(-1) + (-2)*0 = 0; iar în al doilea: ½ * 1 + ½ * (-1) + (-1)*0 = 0. În concluzie, ambele cazuri respectă criteriul ortogonalității.

Contrastele nestandardizate non-ortogonale se aseamănă până la un punct cu cele ortogonale. Totuşi, în cazul lor nu este necesară respectarea celei de-a doua reguli enunțate mai sus, conform căreia, o dată ce o grupă a fost desprinsă pentru a fi comparată cu un set de grupe nu mai poate fi comparată ulterior. În consecință, suma produselor a două contraste nu va mai fi zero în acest caz. Un exemplu de set de contraste nestandardizate non-ortogonale este:

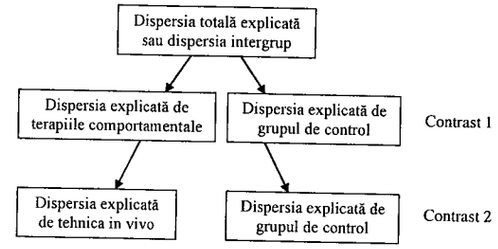

Dispersia totală explicată sau dispersia intergrup
Dispersia explicată de terapiile comportamentale
Dispersia explicată de grupul de control
Contrast 1
Dispersia explicată de tehnica in vivo
Dispersia explicată de grupul de control
Contrast 2

Figura 4.4. *Exemplu de partiționare a dispersiei în cazul contrastelor non-ortogonale*

O asemenea situație ar putea fi transpusă în următorii coeficienți de ponderare:

	„in vivo"	desensibilizare sistematică	control
Contrast 1	1	1	-2
Contrast 2	1	0	-1

În această situație se observă reintroducerea grupului de control în cel de-al doilea contrast după ce în prealabil a fost comparat singur în contrastul 1. Acest fapt nu este permis în contrastele ortogonale, prin urmare nici suma produselor a două contraste nu va mai fi zero [1*1 + 1*0 + (-2)*(-1) = 3].

Dar de ce este important să diferențiem între contrastele ortogonale şi cele non-ortogonale? Primele au marele avantaj de a controla eroarea setului de comparații, fapt ce menține în limite normale valoarea pragului de semnificație. În schimb, contrastele neortogonale, plătesc lipsa de independență dintre comparații, fapt ce conduce la o creştere a riscului de a produce eroarea de tip I. În aceste condiții este necesară foarte multă precauție în interpretarea contrastelor neortogonale, fapt pentru care recomand, de câte ori este posibil, utilizarea exclusivă a contrastelor ortogonale.

Iată o analiză statistică concretă, pe baza exemplului nostru. Ipotezele pe care dorim să le testăm sunt:

a) fobia de vampiri a subiecților cuprinşi în grupele terapeutice comportamentale va fi semnificativ mai mică decât cea întâlnită în grupul de control;

Ho1: 1 μ „in vivo" + 1 μ „desensibilizare" = 2 μ „control"

b) terapia „in vivo" este mai eficientă în acest scop în comparație cu desensibilizarea sistematică.
Ho2: 1 μ „in vivo" = 1 μ „desensibilizare"

Se observă că ambele ipoteze sunt specifice, spre deosebire de o ipoteză generală F de genul: „în ce măsură există o diferență semnificativă între grupele cercetării în reducerea fobiei de vampiri".
Ho: 1 μ „in vivo" = 1 μ „desensibilizare" = 1 μ „control"

În cazul utilizării contrastelor, rezultatele pot fi exprimate în $F_{contrast}$ sau t. Indiferent de modul de exprimare, valorile pot fi transformate uşor, cunoscându-se faptul că:

$$F = t^2 \qquad (4.4)$$

Primul din cele două grade de libertate ale lui $F_{contrast}$ este întotdeauna unu, în timp ce al doilea grad de libertate este egal cu cel al lui t dintr-o anumită situație. Astfel, dacă t = 3, la 12 grade de libertate, atunci F (1, 12) = 9.

Contrast Coefficients

	situatie experimentala		
Contrast	in vivo	desensibilizare sistematica	control
1	1	1	-2
2	1	-1	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
frica de vampiri	Assume equal variances	1	-8,75	,886	-9,874	33	,000
		2	-1,08	,512	-2,117	33	,042
	Does not assume equal variances	1	-8,75	,882	-9,926	22,255	,000
		2	-1,08	,514	-2,106	21,543	,047

Figura 4.5. *Output SPSS 11.0 cu modul de prezentare a rezultatelor contrastelor*
SPSS este o marcă înregistrată.

În primul tabel din output sunt prezentați coeficienții atribuiți celor două contraste, pentru a diferenția între cele două ipoteze specifice prezentate. Al doilea tabel din output prezintă valorile lui t pentru fiecare dintre cele două contraste specificate, inclusiv cu ajustări în cazul în care nu există o egalitate între dispersiile grupurilor comparate. Pentru problema de față, exista egalitatea dispersiei (testul Levene a fost nesemnificativ statistic), prin urmare, primele două rânduri din al doilea tabel al figurii 4.5. sunt cele care ne interesează.

Se observă, în cazul primei ipoteze că, t (33) = 9,87, p < .01, rezultat ce indică o diferență semnificativă între nivelul fobiei de vampiri în grupul de control şi cel din grupele de subiecți supuse terapiei comportamentale. Rezultatul poate fi scris şi sub forma: F(1, 33) = 97,41, p < .01. Urmărind valorile mediilor, vom observa că ambele intervenții terapeutice sunt semnificativ mai bune în reducerea fobiei de vampiri comparativ cu lipsa oricărei intervenții (grupul de control).

În cazul celei de-a doua ipoteze se observă că t (33) = 2,11, p < .05 este, de asemenea, semnificativ statistic. Ținând cont de conținutul celui de-al doilea contrast (tehnica „in vivo" este mai eficientă decât cea bazată pe desensibilizare sistematică) şi urmărind mediile celor două grupe experimentale, vom concluziona că, într-adevăr, o terapie comportamentală bazată pe tehnica „in vivo" este semnificativ mai bună în a reduce fobia de vampiri, în comparație cu o intervenție terapeutică bazată pe desensibilizare sistematică.

Cititorii atenți au observat, probabil, un aspect deosebit. Prin testele post hoc, mai exact prin aplicarea procedurii Tukey, nu au apărut diferențe semnificative între cele două modalități terapeutice, în timp ce contrastele ortogonale au sesizat că terapia „in vivo" este mai eficientă.

Cum explicăm o asemenea situație şi cum procedăm în astfel de cazuri? Explicația constă în faptul că tehnicile a priori (ex. contrastele nestandardizate ortogonale) au o putere mai mare de a detecta diferențele dintre populațiile din care au fost extrase cele două grupe sau setul de grupe. Prin urmare, în asemenea situații, ne vom baza mai mult pe rezultatele obținute în urma aplicării contrastelor ortogonale decât pe cele rezultate din aplicarea unor comparații post hoc. Acesta este un motiv suplimentar, care ar trebui să determine cercetătorii să prefere, ori de câte ori au ocazia, aplicarea comparațiilor a priori, în dauna celor post hoc.

Mai mult, există situații, foarte rare, în care testul F general nu detectează diferențe între grupele cercetării, în timp ce $F_{contrast}$ sau testele t corespondente ar putea indica diferențe între grupe. În asemenea situații, vom proceda într-un mod similar, bazându-ne pe ideea că aceste comparații a priori au o putere mai mare de detectare a diferențelor reale dintre populațiile din care au fost extrase grupele în comparație cu testul general F (Rosenthal, Rosnow şi Rubin, 2000a).

4.1.2.2. Mărimea efectului în cazul contrastelor

Vom aborda în continuare problema mărimii efectului în cazul utilizării contrastelor. Vă reamintim că prin intermediul lor vom putea constata în ce măsură diferențele semnificative observate sunt substanțiale sau nu. Pentru calculul celor două mărimi ale efectului am utilizat aceleaşi formule 4.3. şi 4.4. Astfel, în cazul primului contrast ce viza diferențele dintre cele două intervenții terapeutice şi grupul de control am obținut o mărime a efectului r de 0,72, rezultat care indică o scădere foarte mare a valorilor fobiei de vampiri la persoanele care au fost supuse oricărei intervenții terapeutice dintre cele două amintite. În cazul celui de-al doilea contrast semnificativ statistic r a fost doar 0,03, fapt ce arată că terapia in vivo duce la o scădere semnificativă, dar foarte mică a nivelului fobiei în comparație cu

persoanele supuse desensibilizării sistematice. Altfel spus, deşi cele două grupe diferă semnificativ, din punct de vedere practic nu există diferenţe calitative între cele două grupe de subiecţi, terapia in vivo având un avantaj foarte mic în faţa terapiei bazată pe desensibilizare sistematică.

4.1.2.3. Contrastele standardizate

Această categorie de contraste constă într-un set de comparaţii standardizate. Ele pot fi ortogonale sau non-ortogonale. SPSS-ul dispune, în meniu, de şapte contraste standardizate, dintre care doar ultimele două din tabelul 4.2. sunt ortogonale. Field (2000) oferă o sinteză excelentă a lor. Pentru a simplifica recunoaşterea lor în SPSS am decis păstrarea denumirii lor din engleză.

Tabelul 4.2. Descrierea contrastelor standardizate prezente în SPSS

Nume	Descriere	Contraste	Trei grupe	Patru grupe
Deviation (first)	Compară fiecare grup (cu excepţia primului) cu o combinaţie a tuturor grupelor.	1 2 3	2 vs. (1,2,3) 3 vs. (1,2,3)	2 vs. (1,2,3,4) 3 vs. (1,2,3,4) 4 vs. (1,2,3,4)
Deviation (last)	Compară fiecare grup (cu excepţia ultimului) cu o combinaţie a tuturor grupelor.	1 2 3	1 vs. (1,2,3) 2 vs. (1,2,3)	1 vs. (1,2,3,4) 2 vs. (1,2,3,4) 3 vs. (1,2,3,4)
Simple (first)	Fiecare grupă este comparată cu prima categorie.	1 2 3	1 vs. 2 1 vs. 3	1 vs. 2 1 vs. 3 1 vs. 4
Simple (last)	Fiecare grupă este comparată cu ultima categorie.	1 2 3	1 vs. 3 2 vs. 3	1 vs. 4 2 vs. 4 3 vs. 4
Repeated	Fiecare grupă, cu excepţia primei categorii este comparată cu grupa imediat precedentă.	1 2 3	1 vs. 2 2 vs. 3	1 vs. 2 2 vs. 3 3 vs. 4
Helmert	Fiecare categorie, cu excepţia ultimei, este comparată cu media grupelor ulterioare.	1 2 3	1 vs. (2,3) 2 vs. 3	1 vs. (2,3,4) 2 vs. (3,4) 3 vs. 4
Difference (Helemert invers)	Fiecare categorie, cu excepţia primei, este comparată cu media grupelor precedente.	1 2 3	3 vs. (2,1) 2 vs. 1	4 vs. (3,2,1) 3 vs. (2,1) 2 vs. 1

Dacă am notat grupul „in vivo" cu 1, grupul „desensibilizare sistematică" cu 2 şi grupul de control cu 3, vom observa că alegerea făcută de noi, în cadrul contrastelor nestandardizate, se regăseşte în situaţia „Difference" pentru trei grupe. Dacă primul contrast testează diferenţa în ceea ce priveşte nivelul fobiei de vampiri dintre grupul de control şi cele două grupe experimentale, cel de-al doilea compară

diferenţele dintre cele două grupe experimentale. Aceste contraste standardizate nu sunt disponibile în SPSS, în opţiunea „one-way ANOVA", ci doar în „General Linear Model - Univariate". În schimb, contrastele nestandardizate se regăsesc doar în „one-way ANOVA" şi nu sunt disponibile, fără sintaxă, în „General Linear Model - Univariate".

Toate contrastele care nu sunt ortogonale, indiferent dacă sunt standardizate sau nu, trebuie evitate datorită şanselor crescute de a comite eroarea de tip I. Numai rezultate puternic semnificative (ex. $p < .001$) pot fi interpretate fără probleme, dacă au fost utilizate contraste non-ortogonale. Un alt aspect tehnic, ţine de modul de prezentare a rezultatelor în SPSS. Acesta nu oferă direct valoarea lui t, însă oferă toate datele necesare calculării sale, cât şi pragul de semnificaţie aferent acestuia. Pentru problema noastră, rezultatele ar fi arătat astfel:

Contrast Results (K Matrix)

situatie experimentala Difference Contrast			Dependent Variable frica de vampiri
Level 2 vs. Level 1	Contrast Estimate		1,083
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		1,083
	Std. Error		,512
	Sig.		,042
	95% Confidence Interval for Difference	Lower Bound	4,238E-02
		Upper Bound	2,124
Level 3 vs. Previous	Contrast Estimate		4,375
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		4,375
	Std. Error		,443
	Sig.		,000
	95% Confidence Interval for Difference	Lower Bound	3,474
		Upper Bound	5,276

Figura 4.6. *Rezultate oferite de SPSS 11.0 pentru contrastul Difference*
SPSS este o marcă înregistrată.

Primul contrast prezentat în figura 4.6. este cel referitor la diferenţele existente dintre cele două intervenţii terapeutice de sorginte experimentală (1 – „in vivo"; 2 – „desensibilizare sistematică"). Valoarea lui t este egală cu raportul dintre estimarea contrastului şi eroarea standard a diferenţelor dintre medii. În acest prim caz, t este 1,083/0,512, adică $t = 2,11$, $p < .05$, tehnicile comportamentale „in vivo" fiind mai eficiente decât cele bazate pe desensibilizare sistematică, în reducerea fobiei de vampiri. Se observă că rezultatul este identic cu cel obţinut în secţiunea contrastelor nestandardizate, fiind vorba de aceeaşi situaţie.

Cel de-al doilea contrast testează diferențele dintre grupul de control (3) şi celelalte două grupe experimentale (1 + 2). Rezultatul t = 9,87, p < .01 este identic cu cel obținut prin contrastele nestandardizate, arătând că cele două intervenții comportamentale sunt mai eficiente în reducerea fobiei de vampiri în comparație cu absența oricărei intervenții (grupul de control).

Aşadar, diferența dintre contrastele standardizate şi cele nestandardizate ține doar de sfera de cuprindere. Există o multitudine de combinații şi, implicit, de contraste nestandardizate. Cele mai des întâlnite sunt preluate de multe programe statistice (ex. SPSS) şi standardizate prin anumite denumiri, pentru a uşura munca de introducere a coeficienților aferenți.

4.1.2.4. Contrastele polinomiale

Aceste contraste reprezintă o categorie aparte de contraste standardizate şi ortogonale. Când variabila independentă este de tip ordinal, permițând astfel, o ordonare a modalităților sale, pot fi aplicate contrastele polinomiale, sub forma unei analize direcționale (Hinkle, Wiersma şi Jurs, 1994). Acestea nu numai că oferă informații legate de diferențele existente între anumite grupe, ci şi arată complementar direcția datelor în funcție de treptele variabilei independente.

Procedurile analizei direcționale pornesc de la condiția că treptele variabilei independente sunt egal distanțate, iar numărul de subiecți din fiecare grupă este egal. Ambele condiții pot fi încălcate, însă în acele condiții se impun criterii conservatoare ale pragurilor de semnificație, cum ar fi un p < .01.

Exemplul ales pentru explicarea contrastele anterioare nu este potrivit pentru a apela la contrastele polinomiale, deoarece cele două grupe de terapie comportamentală nu sunt intrinsec ordonabile, fiind mai degrabă variabile de tip nominal. De aceea, vom apela la un alt exemplu:

> Pentru a testa eficiența şedințelor de hipnoterapie pe un eşantion de subiecți bulimici au fost realizate cinci grupe, în care subiecții au fost incluşi aleatoriu. Grupul 1 este păstrat ca grup de control, al doilea grup urmează şedințe de hipnoterapie o dată pe săptămână; al treilea grup face acelaşi lucru, dar de două ori pe săptămână, al patrulea grup are trei şedințe săptămânale, în timp ce al cincilea grup urmează patru şedințe săptămânale de hipnoterapie.

Supunând aceste grupe la o analiză direcțională vom putea întâlni unul dintre cele patru cazuri de evoluție a datelor, prezentate grafic în cele ce urmează:

- direcție liniară (poate apărea în design-uri experimentale cu două sau mai multe grupe);
- direcție curbilinie (poate apărea în design-uri experimentale cu trei sau mai multe grupe);
- direcție cubică (poate apărea în design-uri experimentale cu patru sau mai multe grupe);
- direcție cuartică (poate apărea în design-uri experimentale cu cinci sau mai multe grupe).

Direcția liniară, are o evoluție constantă de la o treaptă la alta.

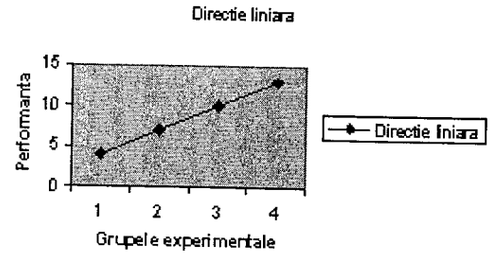


Figura 4.7. *Direcție liniară*

Direcția curbilinie presupune o evoluție constantă până la un anumit punct (ex. ascendentă), după care sensul direcției se schimbă (ex. evoluție descendentă).

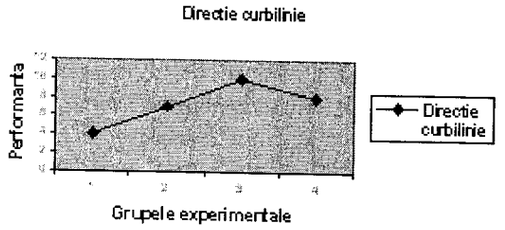


Figura 4.8. *Direcție curbilinie*

Direcția cubică, presupune o evoluție inconstantă, cu două schimbări de direcție pe parcursul trecerii de la o treaptă a variabilei independente la alta.

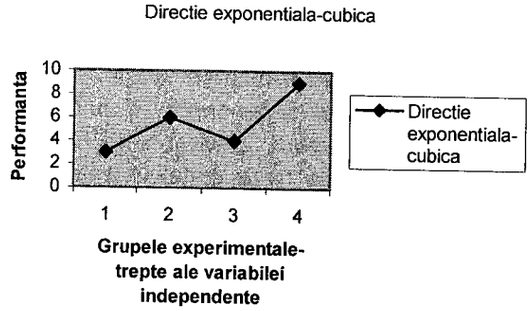


Figura 4.9. *Direcție cubică*

Ultima posibilitate de evoluție a direcției este cuartică. Această direcție este foarte inconsecventă, având trei puncte de schimbare a direcției.

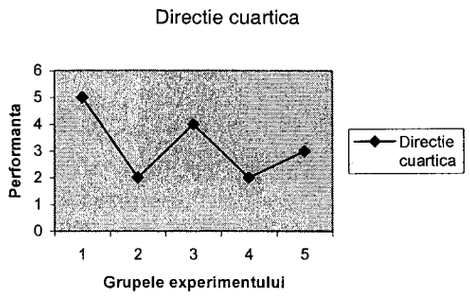


Figura 4.10. *Direcție cuartică (neregulată)*

Contrastele polinomiale nu trebuie specificate în SPSS deoarece au valori fixe, pentru direcții şi număr de grupe diferite. Iată câteva valori ale acestora pentru situațiile experimentale cu trei până la şase grupe (Hinkle, Wiersma, Jurs, 1994).

K = 3 (trei grupe de subiecți)

Liniar	1	0	-1
Curbiliniu	1	-2	1

K = 4 (patru grupe de subiecți)

Liniar	-3	-1	1	3
Curbiliniu	1	-1	-1	1
Cubic	-1	3	-3	1

K = 5 (cinci grupe de subiecți)

Liniar	-2	-1	0	1	2
Curbiliniu	2	-1	-2	-1	2
Cubic	-1	2	0	-2	1
Cuartic	1	-4	6	-4	1

K = 6 (şase grupe de subiecți)

Liniar	-5	-3	-1	1	3	5
Curbliniu	5	-1	-4	-4	-1	5
Cubic	-5	7	4	-4	-7	5
Cuartic	1	-3	2	2	-3	1

Dezavantajul utilizării acestor contraste polinomiale constă în posibilitatea de a obține mai multe valori F semnificative, ceea ce implică mai multe direcții posibile de evoluție a variabilei dependente în funcție de treptele variabilei independente.

Să presupunem că dintre toate testele t calculate, doar t liniar a ieşit semnificativ statistic. Această situație ar fi arătat că direcția de evoluție a lucrurilor este una liniară. Consultând mediile celor cinci grupe vom trage, probabil, următoarele concluzii: şedințele de hipnoterapie se dovedesc a fi eficiente în tratarea bulimiei, cu cât numărul de şedințe de hipnoterapie creşte, cu atât scad semnificativ manifestările bulimice.

4.2. ANOVA factorială

Până în prezent am discutat cazurile de ANOVA în care era prezentă o singură variabilă independentă cu mai multe trepte şi o singură variabilă dependentă. Aceasta nu este însă şi singura modalitate de a realiza o cercetare. Design-ul acesteia poate include mai multe variabile independente. În acest caz se utilizează ANOVA factorială, prin care se urmăreşte studiul influenței simultane a două sau mai multe variabile independente asupra unei variabile dependente.

Există mai multe avantaje în utilizarea unui design experimental factorial. Un prim câştig este dat de un control eficient al unor variabile de interes, care în ANOVA simplă sunt excluse din studiu. Un alt avantaj constă în eficiența studiului, deoarece două sau mai multe variabile independente sunt analizate simultan, fiind extrase informații referitoare atât la fiecare factor în parte, cât şi la interacțiunea dintre aceştia, vizând influența comună asupra VD.

Din această perspectivă, ANOVA factorială pentru două variabile independente conduce la calcularea a trei teste statistice F. Două dintre ele se realizează pentru a găsi efectele principale (efectul principal al variabilei A şi efectul principal al variabilei B), iar cel de-al treilea test măsoară interacțiunea dintre ele. Efectele principale sunt analoge dispersiei intergrup din ANOVA simplă, iar testul interacțiunii (simbolizat AB) ne arată dacă rezultatele obținute la diferite nivele ale variabilei A sunt dependente de nivelul factorului B şi viceversa.

În cele ce urmează ne vom opri doar asupra testului ANOVA bifactorial (2 variabile independente cu mai multe trepte fiecare), fiind însă posibile şi experimente care necesită mai mult de doi factori. Algoritmul de rezolvare a acestor situații este asemănător cu ANOVA bifactorial. De exemplu, un design factorial complet cu trei variabile independente ar duce la calcularea a şapte teste F generale. Trei dintre ele ar urmări efectele principale ale celor trei variabile independente A, B, C, alte trei teste F ar urmări interacțiunea dintre fiecare pereche de variabile (AB, AC, BC), iar ultimul F ar surprinde o interacțiune complexă a celor trei variabile studiate (ABC).

Orice ANOVA factorial are specificat designul experimental realizat sub forma unor cifre. Un design factorial cu două variabile independente, fiecare având câte trei trepte, va fi notat ANOVA 3x3. Un alt design, în care prima variabilă are două trepte, iar cea de a doua cinci, va fi simbolizat ANOVA 2x5. Un design factorial, cu trei variabile independente, în care primele două variabile au câte 3 trepte, iar cea de-a treia doar două, va fi notat ANOVA 3x3x2 ş.a.m.d.

Voi prezenta în continuare forma grafică a unui design experimental bifactorial, în care cele două variabile independente au câte trei trepte (tabelul 4.3).

După cum se observă, din combinarea celor doi factori, fiecare cu trei trepte, se obțin nouă situații experimentale diferite. Acestea poartă numele de **celule experimentale.**

Tabelul 4.3. Exemplu de design experimental de tip 3 x 3

	B1	B2	B3	
A1	**A1B1**	**A1B2**	**A1B3**	Medie A1
A2	**A2B1**	**A2B2**	**A2B3**	Medie A2
A3	**A3B1**	**A3B2**	**A3B3**	Medie A3
	Medie B1	Medie B2	Medie B3	

Dacă în cazul ANOVA simplă dispersia totală era suma dispersiei intergrup cu dispersia intragrup, în cazul ANOVA bifactorială există patru surse ale dispersiei totale, şi anume: dispersia intracelulă (echivalentă cu dispersia intragrup); dispersia de-a lungul mediei liniilor (datorată factorului A); dispersia de-a lungul mediei coloanelor (datorată factorului B) şi dispersia datorată interacţiunii dintre cele două variabile independente (AB). Suma ultimelor trei dispersii este echivalentul dispersiei intergrup.

De aceea, în cazul ANOVA bifactorial sunt testate trei ipoteze nule. Primele două se referă la efectul variabilelor independente luate separat ($\mu A1 = \mu A2 = \mu A3$, respectiv $\mu B1 = \mu B2 = \mu B3$), iar cea de-a treia se referă la interacţiunea dintre cei doi factori.

Pentru a concretiza elementele teoretice prezentate în cadrul ANOVA factorială, voi oferi drept model de rezolvare o problemă tipică vizând memoria.

Pornind de la experimentele clasice ale lui Loftus şi Palmer (Eysenck şi Keane, 2000) care au arătat că memoria are o natură reconstructivă, un cercetător doreşte să studieze în ce măsură relatările martorilor oculari sunt influenţate de întrebările adresate şi de experienţa faţă de situaţia în cauză, având drept obiectiv stabilirea gradului de încredere în depoziţiile martorilor la procese.

Experimentul descris în continuare este o adaptare a cercetărilor iniţiale realizate de autorii menţionaţi. Aceştia propun subiecţilor să vizioneze un film despre un accident de circulaţie. După vizionare subiecţilor li se cere să estimeze viteza cu care rulau maşinile în momentul accidentului (variabila dependentă), cercetătorii fiind interesaţi de influenţa a două variabile independente.

O primă variabilă manipulată de experimentatori este modul de adresare a întrebării. Astfel, ei cer unui grup să estimeze viteza cu care s-au "zdrobit" maşinile; unui al doilea grup să estimeze viteza cu care "s-au lovit" maşinile, în timp ce în a treia situaţie, subiecţii au fost întrebaţi ce viteză aveau maşinile în momentul accidentului. Cea de a doua variabilă independentă a fost experienţa subiecţilor care a avut două trepte: şoferi şi ne-şoferi. Aşadar, avem un model ANOVA factorial de tip 3x2 (trei linii şi două coloane).

Ipoteza cercetătorului este că relatările subiecţilor sunt influenţate atât de expresiile utilizate (întrebările anchetatorului), cât şi de experienţa subiecţilor în situaţia descrisă (profesiunea martorului) în sensul că subiecţii – şoferi pot estima mai bine decât non-şoferii viteza de deplasare a automobilelor).

Baza de date aferentă exemplului este anova factoriala.sav, iar calea de obţinere a rezultatelor în SPSS este „Analyze" - „General Linear Model" - „Univariate".

Tests of Between-Subjects Effects

Dependent Variable: viteza masinilor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1702,800[a]	5	340,560	7,277	,000
Intercept	38880,000	1	38880,000	830,769	,000
SUGESTII	1139,400	2	569,700	12,173	,000
EXPERIEN	97,200	1	97,200	2,077	,162
SUGESTII * EXPERIEN	466,200	2	233,100	4,981	,016
Error	1123,200	24	46,800		
Total	41706,000	30			
Corrected Total	2826,000	29			

a. R Squared = ,603 (Adjusted R Squared = ,520)

Figura 4.11. *Output SPSS 11.0 cu tabelul sumar ANOVA factorială*
SPSS este o marcă înregistrată.

Eliminând elementele care nu ne interesează în acest moment, din informaţiile prezentate în modelul general liniar din SPSS, obţinem următorul tabel ANOVA:

Tabelul 4.4. Tabelul sumar ANOVA 3x2 factorial

Sursa dispersiei	SS	df	MS	F	p
Linii (A) - sugestii	1139,4	2	569,7	**12,17**	<.01
Coloane (B) – experienţă	97,2	1	97,2	2,07	>.10
Interacţiune AxB	466,2	2	233,1	**4,98**	<.05
Intracelulă	1123,2	24	46,8		
Totală	2826,0	29			

Aşadar, se observă că din cele trei raporturi F, două sunt semnificative statistic: F „sugestii" [$F(2, 24) = 12,17$, $p < .01$] şi F interacţiune [$F(2, 24) = 4,98$, $p < .05$]. Primul rezultat indică un efect principal semnificativ, care poate fi transpus în următoarea afirmaţie: estimarea vitezei de deplasare a maşinilor în momentul accidentului este influenţată semnificativ de modul de adresare a întrebării, prin cuvintele alese în acest scop. Aceasta înseamnă că, dacă ignorăm experienţa celor două categorii de subiecţi (şoferi vs. non-şoferi), există o diferenţă semnificativă în estimarea vitezei, influenţată de modul de formulare a întrebării. Fiind un test F general, de tip omnibuz, nu putem deduce direct cum evoluează lucrurile. Spre exemplu, utilizarea termenului „zdrobit" conduce la estimarea unei viteze mai mari decât folosirea termenului „lovit"? Există diferenţe semnificative

între utilizarea termenului „zdrobit" față de utilizarea sintagmei „cu ce viteză circulau maşinile în momentul accidentului?" etc. Pentru a observa care diferențe sunt semnificative putem apela, fie la contraste, fie la comparații post hoc, ambele metode fiind deja prezentate în secțiunea destinată analizei dispersionale simple.

În urma aplicării testului post hoc Tukey, au fost observate diferențe semnificative în două dintre cele trei comparații. Mai precis, ignorând aspectul experienței în conducerea autovehiculelor, persoanele cărora li s-a adresat întrebarea incluzând termenul „zdrobit" au estimat viteze de deplasare semnificativ mai mari în comparație cu situația utilizării termenului „lovit" sau în comparație cu grupul asupra căruia s-a folosit expresia neutră „viteza cu care circulau maşinile". În schimb, nu au fost sesizate diferențe semnificative în ceea ce priveşte viteza de deplasare a automobilelor în cazul utilizării expresiei „lovit" față de cazul exprimării „viteza cu care circulau maşinile".

Cel de-al doilea F din tabelul 4.4. a testat în ce măsură există o diferență semnificativă între şoferi şi non-şoferi, indiferent de cuvintele utilizate, referitoare la estimarea vitezei cu care circulau maşinile în momentul accidentului. Rezultatul, $F(1, 24) = 2{,}07$, $p > .10$, a fost nesemnificativ statistic, astfel că nu s-au găsit diferențe semnificative între cele două categorii de persoane cu privire la viteza de deplasare a maşinilor. Aceasta înseamnă că, ignorând tipul de întrebări adresate, nu există diferențe între şoferi şi non-şoferi în estimarea vitezei de deplasare a maşinilor, în momentul accidentului din filmul respectiv.

În schimb, cel de-al treilea test F, vizând gradul de interacțiune dintre cei doi factori, este semnificativ statistic, existând o asociere între treptele celor două variabile independente. Cu alte cuvinte, efectul modului de formulare a întrebărilor asupra estimării vitezei de deplasare a maşinilor a fost diferit pentru şoferi față de non-şoferi. Cum a fost diferit, este o altă poveste, ce necesită înțelegerea conceptului de efect simplu.

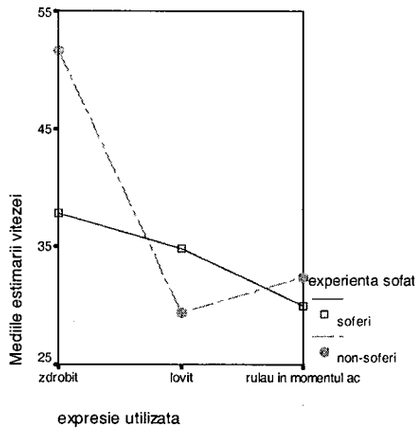


Figura 4.12. *Reprezentarea grafică în SPSS 11.0 a interacțiunii dintre cei doi factori*

SPSS este o marcă înregistrată.

Un mod alternativ, dar simplificat, de a observa această evoluție, se obține prin reprezentarea grafică a interacțiunii dintre cei doi factori. Acesta se poate obține direct din SPSS utilizând aceeaşi cale de obținere a rezultatelor în cazul ANOVA factorial, „Analyze" – „General Linear Model" – „Univariate", după care se selectează butonul – „Plots". Înainte de a face acest lucru trebuie subliniată următoarea idee: deşi efectele principale au rolul lor, dacă interacțiunea dintre doi factori este semnificativă, aceasta devine primordială în interpretarea statistică şi psihologică a rezultatelor unei cercetări.

Ca o regulă generală, în cele mai multe situații când cele două linii se intersectează, precum în figura 4.12., există o interacțiune semnificativă între acei doi factori. Din desen, se pot desprinde câteva observații. Diferențele cele mai mari, dintre şoferi şi non-şoferi, se observă în cazul în care experimentatorul a folosit sintagma „cu ce viteză s-au zdrobit maşinile în accident". Astfel, non-şoferii tind să fie influențați mai puternic de această expresie, acordând o viteză de deplasare mult mai mare decât media estimărilor făcute de şoferi. De asemenea, putem nota că cea mai mică diferență dintre cele două categorii de persoane se observă în situația în care experimentatorul adoptă o atitudine neutră, utilizând expresia „Cu ce viteză circulau maşinile în momentul accidentului?". Ar fi bine ca aceste observații grafice să fie completate de analiza efectelor simple, o metodă riguroasă, dar din nefericire absentă din meniul SPSS-ului, ea fiind disponibilă doar prin apelul la sintaxă.

4.2.1. Metoda efectelor simple

Tehnica are un caracter post hoc şi este menită să surprindă în detaliu diferențele dintre anumite celule experimentale. Ea se deosebeşte de metodele post hoc obişnuite, fiind mult mai analitică (Keppel, 1991; Pedhazur şi Schmelkin, 1991). Spre exemplu, prin metoda Tukey se observă dacă există diferențe între situațiile experimentale determinate de modul de formulare a întrebărilor, ignorând aspectul experienței în conducerea automobilului. De exemplu, vom putea compara dacă există diferențe semnificative între persoanele întrebate „cu ce viteză s-au zdrobit maşinile", respectiv persoanele aflate în grupa în care întrebarea a fost „cu ce viteză s-au lovit maşinile".

În schimb, prin metoda efectelor simple, vom face un pas în plus, determinând posibile diferențe între toate cele şase celule experimentale ale problemei date. De exemplu, vom putea stabili dacă există diferențe semnificative între şoferii care au fost întrebați cu ce viteză circula maşina în momentul accidentului şi non-şoferii care au fost întrebați cu ce viteză s-au zdrobit maşinile.

Procedura de a obține efectele simple în SPSS este extrem de complicată, presupunând utilizarea sintaxelor. De aceea, vom prezenta pe scurt formulele matematice ale efectelor simple. Cei interesați pot consulta un material complet pe site-ul de statistică socială (Sava, 2002).

Pentru a afla efectul simplu în cadrul acestei trepte vom parcurge trei paşi:

1. vom calcula suma pătratelor aferentă acestui efect simplu;
2. pe baza sumei pătratelor şi cunoscând gradele de libertate, vom obține media pătratică, iar din aceasta testul F aferent;
3. pe baza lui F vom stabili dacă treapta respectivă contribuie semnificativ la interacțiunea dintre cei doi factori.

Vom presupune că au fost obținute următoarele rezultate în situația experimentală „cu ce viteză s-au zdrobit maşinile".

Tabelul 4.5. Rezultatele obținute de şoferi şi non-şoferi în situația experimentală „zdrobit"

Factorul 1 „mod de formulare a întrebărilor"	Şoferi (B1)		Non-şoferi (B2)		TOTAL şoferi şi non-şoferi
	X	X^2	X	X^2	
treapta 1 - „zdrobit"	51	2601	63	3969	Σ treapta 1 = 189 + 258 = 447
	45	2025	57	3249	
	39	1521	51	2601	Σ (treapta 1)² = 7623 + 13644 = 21267
	30	900	48	2304	
	24	576	39	1521	
Sumă (Σ)	189	7623	258	13644	
Viteza medie	37,8		51,6		44,7

Formula de calcul a sumei pătratelor pentru fiecare treaptă este:

$$SS_l = \frac{\left(\sum_{c1l} X\right)^2}{n_{c11}} + \ldots + \frac{\left(\left(\sum_{cnl} X\right)^2\right)}{n_{cn1}} - \frac{\left(\sum_{totl} X\right)^2}{n_{tot1}} (4.5)$$

unde: SSl este suma pătratelor unei trepte;
nc1l este numărul de subiecți din prima celulă a unei trepte;
ncnl este numărul de subiecți din ultima celulă a unei trepte;
ntotl este numărul de subiecți ai unei trepte indiferent de celulă;
Σc1lX este suma lui X din prima celulă a unei trepte;
ΣcnlX este suma lui X din ultima celulă a unei trepte;
ΣtotlX este suma tuturor X din treapta respectivă.

În problema de față avem doar două celule ale treptei „zdrobit" reprezentate de grupul de subiecți şoferi, respectiv grupul de subiecți non-şoferi.

Astfel, vom obține:

$$SS_{A1} = \frac{189^2}{5} + \frac{258^2}{5} - \frac{447^2}{10} = 476{,}1$$

Pasul al doilea constă în obținerea lui F – efect simplu – pentru treapta „zdrobit" a variabilei „mod de formulare a întrebărilor". Acesta se obține simplu, pe baza formulei:

$$F = \frac{MS_{efect\_simplu}}{MS_{\text{int}\, racelula}} (4.6)$$

Dacă media pătratică intracelulă există în tabelul sumar ANOVA, lipseşte media pătratică a efectului simplu. Aceasta este egală cu raportul dintre suma pătratelor efectului respectiv şi gradul de libertate al celeilalte variabile cu care interacționează variabila în cauză. În cazul de față, media pătratică calculată a acestui efect simplu este egală cu suma pătratelor calculate, adică 476,1, deoarece gradul de libertate a variabilei „experiență în şofat" este unu. Înlocuind în formula de mai sus, vom obține un F - efect simplu - pentru treapta „zdrobit" a variabilei „modul de formulare a întrebării" de 10,17 (476,1:46,8). La fel se procedează şi în cazul celorlalte trepte ale variabilelor incluse în studiu.

Reamintim datele din tabelul sumar ANOVA pentru a clarifica toate valorile introduse în calculele de mai sus:

Tabelul 4.6. Tabelul sumar ANOVA pentru exemplul dat

Sursa dispersiei	SS	df	MS	F	p
Formularea întrebărilor (A)	1139,4	2	569,7	**12,17**	<.01
Experiență şofat (B)	97,2	1	97,2	2,07	>.05
Interacțiune	466,2	2	233,1	**4,98**	<.05
Intracelulă (Reziduală)	1123,2	24	46,8		
Totală	2826	29			

În urma calculării celor cinci efecte simple, corespondente fiecărei trepte a variabilelor independente, după modelul explicat anterior, vom obține un tabel sumar ANOVA nuanțat.

Tabelul 4.7. Tabelul ANOVA extins, incluzând valorile F pentru efectele simple

Sursa dispersiei	SS	df	MS	F	p
Formularea întrebării (A)	1139,4	2	569,7	**12,17**	<.01
B1 – şoferi	154,8	2	77,4	1,65	
B2 – non-şoferi	1450,8	2	725,4	**15,50**	<.01
Experiența în conducerea maşinii (B)	97,2	1	97,2	1,24	
A1 – cu ce viteză s-au zdrobit	476,1	1	476,1	**10,17**	<.01
A2 – cu ce viteză s-au lovit	72,9	1	72,9	1,55	
A3 – cu ce viteză circulau	14,4	1	14,4	0,30	
Interacțiune	466,2	2	233,1	**4,98**	<.05
Intracelulă	1123,2	24	46,8		
Totală	2826,0	29			

După ce am aflat valorile lui F – efect simplu (vezi tabelul 4.7.) – pentru fiecare din cele cinci trepte, urmează ultimul pas de a stabili dacă valorile obținute sunt semnificative statistic. Trebuie să reținem că nu vom utiliza pragul de semnificație clasic de .05. În acest caz intervine o modificare a valorii alfa pentru a păstra constantă eroarea setului de comparații. Astfel, în cazul variabilei cu trei trepte „formularea întrebării" vom avea ca limită p = .017 (.05 / 3), în timp ce pentru variabila „experiență în conducerea mașinii" vom avea p =.025 (.05 / 2), deoarece variabila are două trepte: șoferi și non-șoferi.

Din cele cinci efecte simple, două se dovedesc a fi semnificative statistic. Este vorba de interacțiunea dintre formularea întrebării și non-șoferi (A-B2), respectiv dintre experiența în conducere și folosirea sintagmei „zdrobit" (B-A1). Interpretarea acestor efecte simple semnificative se face pe baza observării mediilor situațiilor respective și a graficului de reprezentare a interacțiunii dintre factori. Pentru problema noastră se desprind două concluzii complementare, de nuanțare:

(1) primul efect simplu semnificativ (A-B2) indică o diferențiere mai clară, în aprecierea vitezei de deplasare a mașinii în momentul accidentului, la persoanele care nu posedă carnet de conducere. Acest fapt se datorează probabil lipsei de experiență în aprecierea vitezelor, determinând această categorie de subiecți să fie mai ușor influențabilă de sugestiile experimentatorului, prin modul de formulare a întrebărilor. Acest aspect se poate observa și în graficul de interacțiune a factorilor, rezultatele subiecților fiind mai variabile decât cele ale șoferilor.

(2) al doilea efect simplu semnificativ (B-A1) indică o diferență între șoferi și non-șoferi în cazul primei situații experimentale. Astfel, sintagma „zdrobit" influențează într-o mai mică măsură șoferii decât non-șoferii. Aceștia din urmă estimează o viteză de deplasare a mașinilor mult mai mare în cazul în care au fost utilizate cuvinte sugestive pentru forța impactului („zdrobit"). În schimb, șoferii sunt mai puțin afectați de acest aspect. Graficul interacțiunii identifică această situație, în care formularea „zdrobit" conduce la cea mai mare diferențiere între șoferi și non-șoferi, în ceea ce privește estimarea vitezei de deplasare a mașinii în momentul producerii accidentului.

4.2.2. Mărimea efectelor în cazul ANOVA factorială

Mărimea efectului poate fi calculată și în cadrul unui design factorial. Prin intermediul acestui indicator vom putea observa în ce măsură diferențele semnificative observate au o anumită importanță practică. Modul de calcul este similar celui prezentat în cazul ANOVA simplă, cu deosebirea că F reprezintă valoarea testului pentru un anumit factor, iar F contrast este valoarea obținută în cadrul unei comparații post hoc sau a unui contrast în cadrul factorului respectiv. În cazul în care un factor are doar două trepte, F contrast este egal cu F al factorului respectiv (formula 4.2.). În cazul unei VI cu trei sau mai multe modalități se

folosește formula 4.3, unde df intergrup devine df asociat factorului respectiv, iar df intragrup devine df corespunzător dispersiei intracelulă.

Astfel, utilizarea expresiei „viteza cu care s-au zdrobit mașinile" a avut un impact puternic asupra estimării vitezei. Efectul înregistrat a fost foarte mare comparativ cu viteza estimată de persoanele care au fost invitate să precizeze viteza mașinilor în momentul lovirii lor, r fiind 0,59, r^2 fiind 0,35. Pentru a ajunge la aceste valori am aplicat formulele 4.3 și 4.4, utilizând valoarea lui F sugestii din tabelul 4.4 și calculând t Tukey pentru comparația dintre grupele „zdrobit" și „lovit". Cu alte cuvinte, 35% din dispersia vitezei estimate de către subiecții investigați este influențată de modul de formulare a întrebării, în cazul sintagmei „viteza cu care s-au zdrobit mașinile". Acest aspect poate avea implicații profunde în veridicitatea depozițiilor martorilor în actul de justiție. Se observă că modul de formulare a întrebării poate induce răspunsuri cu un impact profund diferit în evaluarea unor situații în care lipsesc reperele de apreciere.

4.2.3. Condiții necesare pentru aplicarea ANOVA simplă și ANOVA factorială

(1) Eșantionul a fost selectat randomizat din populație. În cazul în care acest fapt este greu de realizat, se recomandă ca măsură compensatorie distribuirea aleatoare a subiecților în grupele experimentale.
(2) Variabila dependentă este distribuită normal. Această condiție este greu de îndeplinit în cazul în care grupele experimentale sunt constituite dintr-un număr mic de subiecți (ex. n = 10).
(3) Fiecare persoană din eșantion este testată o singură dată, astfel încât evaluările să fie independente. Când acest aspect nu este respectat, vom utiliza, foarte probabil, ANOVA cu măsurători repetate.
(4) Dispersia subiecților împărțiți pe grupe experimentale să fie egală. Testul Levene este special construit pentru a observa în ce măsură este realizată această omogenitate a dispersiei.

Tehnicile ANOVA sunt destul de robuste la nerespectarea primelor două condiții. Cu alte cuvinte, indicele general F nu este afectat foarte mult de primele două condiții (Hinkle, Wiersma și Jurs, 1994). În schimb, neîndeplinirea condiției patru, absența unei omogenități a dispersiei dintre grupe, afectează semnificativ valorile lui F. Într-o asemenea situație sunt posibile două măsuri alternative: (a) renunțarea la aplicarea tehnicilor ANOVA și utilizarea unor tehnici corespondente neparametrice; (b) aplicarea tehnicilor parametrice ANOVA prin apelul la o corecție a gradelor de libertate de citire a lui F din tabelul distribuției omonime. Tabelul poate fi consultat în aproape orice manual introductiv de statistică.

În situația din urmă, valoarea lui F obținută va fi comparată cu valoarea prezentată în tabelul lui F pentru 1 și N-1 grade de libertate și nu cu valoarea sa corespondentă de la k-1 și N-k-1 grade de libertate, dacă omogenitatea dispersiei

era îndeplinită. Menționăm că N este numărul total de subiecți din eşantion, iar k este numărul de grupe sau de celule experimentale din cercetarea respectivă.

4.3. ANOVA cu măsurători repetate

În condițiile în care aceiaşi subiecți sunt testați de mai multe ori vom apela la ANOVA cu măsurători repetate. Dacă ANOVA simplă era corespondentul testului t pentru două eşantioane independente, într-o anumită măsură ANOVA cu măsurători repetate este corespondentul testului t pentru două eşantioane perechi. Am utilizat sintagma „într-o anumită măsură", deoarece am putea vorbi de ANOVA cu măsurători repetate, atât în cazul în care un singur factor este luat în considerare, cât şi în situația în care mai mulți factori sunt măsurați repetat (ANOVA factorială cu măsurători repetate). Termenul de măsurători repetate implică evaluarea unui individ de două sau mai multe ori privitor la aceeaşi variabilă dependentă.

Vom utiliza un exemplu de ANOVA cu măsurători repetate apelând la un exemplu din psihologia socială. Tajfel recunoaşte categorizarea socială ca o componentă de bază a identității sociale. Acest proces cognitiv de structurare a stimulilor din mediul înconjurător pentru a simplifica percepția nu se rezumă doar la mediul social. Vom adapta un studiu descris pe scurt de Lyons (1998) şi aparținând lui Tajfel şi Wilkes. Autorii au demonstrat tendința de accentuare a diferențelor în cazul unei categorizări prealabile a stimulilor. Concret, subiecții au fost rugați să evalueze, în număr de cuvinte, lungimea unor replici dintr-o piesă de teatru. Din cele opt replici selectate, patru erau mai scurte, iar patru erau mai lungi. Vom prezenta trei situații experimentale, dintre care primele două respectă studiul inițial, cea de a treia fiind introdusă de noi cu scopul de a utiliza ANOVA cu măsurători repetate în dauna unui simplu test t dependent. Prima situație experimentală consta în faptul că cele patru replici mai lungi erau etichetate cu litera A, iar celelalte patru replici mai scurte erau identificate cu B. A doua situație a studiului inițial, poate fi etichetată ca grup de control, deoarece cele opt replici nu erau asociate cu nici o literă. Cea de-a treia situație, introdusă de noi, constă în asocierea mixtă a literelor A şi B cu cele opt replici, astfel încât să existe două replici mai lungi şi două replici mai scurte asociate literei A, iar restul replicilor (două replici mai lungi şi două replici mai scurte) fiind etichetate de litera B.

Conform efectului de accentuare activat prin etichetarea stimulilor, subiecții au exagerat diferențele dintre replicile mai lungi şi cele mai scurte în cazul etichetării primelor cu litera A şi a celorlalte cu litera B, comparativ cu lipsa etichetării acestora în grupa de control. Pornind de la aceste premise, ne vom aştepta ca etichetarea replicilor din a treia situație experimentală să reducă diferențele dintre replicile mai lungi şi cele mai scurte, datorită identificării mixte a acestora cu litera A, respectiv B.

Pentru a fi un design cu măsurători repetate, merită să precizăm că au fost folosiți aceiaşi subiecți în cele trei situații experimentale diferite. În plus, pentru a elimina aspectele de învățare, au fost folosite trei forme paralele ale replicilor, care aveau acelaşi număr de cuvinte fiecare, iar pentru a elimina posibilele efecte de ordine sau oboseală, prezentarea replicilor a fost contrabalansată, astfel încât unii subiecți au răspuns mai întâi la prima situație experimentală, alții la situația de control, iar alții la ultima situație experimentală.

Pe baza unor date imaginare vom testa următoarea ipoteză generală. Categorizarea stimulilor afectează modul de percepție a acestora. Cele două ipoteze specifice sunt:

a) includerea în aceeaşi categorie a stimulilor similari conduce la accentuarea percepției deosebirilor dintre stimulii diferiți în comparație cu absența categorizării. Ipoteza nulă pentru această situație este $\mu 1 = \mu 2$, adică nu există diferențe semnificative în aprecierea distanței dintre replicile mai lungi şi cele mai scurte în primele două situații;
b) includerea în aceeaşi categorie a stimulilor diferiți va conduce la reducerea percepției deosebirilor dintre ei în comparație cu absența categorizării. Ipoteza nulă este în acest caz $\mu 2 = \mu 3$, adică nu există diferențe semnificative între aprecierea distanței dintre replicile mai lungi şi cele mai scurte în situația ultimelor două grupe.

Altfel spus, utilizarea aceleiaşi etichete pentru stimuli similari conduce la creşterea asemănărilor percepute dintre ei, în timp ce utilizarea aceleiaşi etichete pentru stimuli diferiți conduce la scăderea deosebirilor percepute dintre ei.

Exemplul prezentat poate fi exersat utilizând fişierul anova repetate.sav, urmând calea „Analyze" - „General Linear Model" - „Repeated Measures". În fereastra deschisă va trebui asociat numărul 3 termenului „fac_1", pentru a defini cele trei situații experimentale întâlnite, prin acționarea butonului „define". În cazul unei ANOVA factorială cu măsurători repetate, înainte de a acționa butonul „define", se apasă „add" pentru a permite definirea numărului de repetări în cazul celui de-al doilea factor.

ANOVA cu măsurători repetate poate fi realizată atât prin testarea lui F general, urmată de comparații post hoc, cât şi prin testarea unor F specifici, prin aplicarea unor contraste. În cazul de față ipotezele impun cea de-a doua abordare, datorită specificității existente în formularea lor. Totuşi, în scop didactic, vor fi analizate rezultatele obținute şi prin metodele post hoc.

Modul de prezentare a tabelului sumar pentru ANOVA cu măsurători repetate din SPSS este puțin diferit de modul de prezentare al tabelului sumar din celelalte cazuri. El este constituit doar din două surse ale dispersiei: cea determinată de factorul cercetării (în cazul nostru, tipul de situație experimentală), respectiv dispersia datorată erorii. De fapt, aceste surse sunt şi cele mai importante pentru calcularea lui F.

Totuşi, această modalitate de prezentare a rezultatelor nu ține seama de specificul tehnicii ANOVA cu măsurători repetate. Spre deosebire de ANOVA

simplă, ce consideră diferențele interindividuale ca parte a dispersiei erorii (neexplicate), în ANOVA cu măsurători repetate putem distinge trei surse ale dispersiei: dispersia explicată (datorată treptelor variabilei independente), dispersia subiecților (datorată diferențelor dintre indivizi) şi dispersia neexplicată, reziduală sau dispersia erorii. Dacă vă reamintiți, F nu era altceva decât raportul dintre dispersia explicată şi cea neexplicată. Din acest punct de vedere, datorită reducerii dispersiei neexplicate, prin eliminarea dispersiei datorate diferențelor individuale, F are mai mari şanse să fie semnificativ statistic. Astfel, un design cu măsurători repetate are o putere statistică mai mare de a detecta situațiile în care variabilele independente influențează o anumită variabilă dependentă.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SITUATIE	Sphericity Assumed	84,500	2	42,250	37,430	,000
	Greenhouse-Geisser	84,500	1,784	47,367	37,430	,000
	Huynh-Feldt	84,500	2,000	42,250	37,430	,000
	Lower-bound	84,500	1,000	84,500	37,430	,000
Error(SITUATIE)	Sphericity Assumed	24,833	22	1,129		
	Greenhouse-Geisser	24,833	19,623	1,266		
	Huynh-Feldt	24,833	22,000	1,129		
	Lower-bound	24,833	11,000	2,258		

Figura 4.13. *Output SPSS 11.0 cu tabelul sumar ANOVA în cazul unui design cu măsurători repetate*
SPSS este o marcă înregistrată.

Pentru a identifica în SPSS dispersia datorată indivizilor şi implicit totalul sumei pătratelor, putem să tratăm un design cu măsurători repetate ca un design normal, introducând un factor aleatoriu denumit „subiecți", pentru a identifica scorurile aparținând aceleiaşi persoane. Tabelul sumar din SPSS devine:

Tests of Between-Subjects Effects

Dependent Variable: diferenta replici lungi - scurte

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	10302,250	1	10302,250	719,903	,000
	Error	157,417	11	14,311[a]		
SITUATIE	Hypothesis	84,500	2	42,250	37,430	,000
	Error	24,833	22	1,129[b]		
SUBJECT	Hypothesis	157,417	11	14,311	12,678	,000
	Error	24,833	22	1,129[b]		
SITUATIE * SUBJECT	Hypothesis	24,833	22	1,129	,	,
	Error	,000	0	,[c]		

a. MS(SUBJECT)
b. MS(GRUPA * SUBJECT)
c. MS(Error)

Figura 4.14. *Output SPSS 11.0 cu tabelul sumar ANOVA factorial, în cazul tratării subiecților ca factori aleatori*
SPSS este o marcă înregistrată.

Se observă, din compararea figurilor 4.13. şi 4.14., că dispersia erorii în cazul măsurătorilor repetate s-a redus la 24,833. Dacă nu am fi eliminat din cadrul ei sursa dispersiei datorate diferențelor dintre indivizi, am fi obținut o valoarea mai mare: 182,250 (157,417 + 24,833). Valoarea ar fi obținută şi dacă am face abstracție de rezultatele care provin de la aceiaşi subiecți (vezi figura 4.15.).

ANOVA

Diferenta replici lungi - scurte

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	84,500	2	42,250	7,650	,002
Within Groups	182,250	33	5,523		
Total	266,750	35			

Figura 4.15. *Output SPSS 11.0 cu tabelul sumar ANOVA simplă pentru problema categorizării*
SPSS este o marcă înregistrată.

Trecând peste aceste aspecte tehnice, se observă că testul F în urma aplicării ANOVA cu măsurători repetate (figura 4.13.) a fost semnificativ statistic [$F(2, 22) = 37.43$, $p < .01$]. Acest rezultat indică respingerea ipotezei nule conform căreia diferențele existente dintre mediile celor trei grupe s-ar datora hazardului. În consecință, etichetarea unor stimuli are un efect semnificativ asupra modului de percepție a gradului de asemănare sau deosebire dintre aceştia.

În consecință, pentru a testa cele două ipoteze specifice enunțate vom apela la contraste. În cazul ANOVA factorială şi ANOVA cu măsurători repetate, SPSS-ul oferă prin meniu doar oportunitatea de a alege între contraste standardizate. Din tabelul 4.2., putem observa că cea mai potrivită alegere pentru testarea celor două ipoteze este contrastul repetat („repeated"). Acesta testează, pe rând, diferența dintre grupele 1 şi 2, respectiv dintre grupele 2 şi 3. Primul contrast va testa prima ipoteză, iar cel de-al doilea, pe cea de a doua. Rezultatele în SPSS sunt prezentate astfel:

Descriptive Statistics

	Mean	Std. Deviation	N
efect de accentuare	18,75	2,491	12
grup de control	17,00	2,174	12
efect de reducere	15,00	2,374	12

Figura 4.16. *Output SPSS 11.0 cu valorile descriptive pentru cele trei grupe oferite de SPSS*
SPSS este o marcă înregistrată.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SITUATIE	Type III Sum of Squares	df	Mean Square	F	Sig.
SITUATIE	Level 1 vs. Level 2	36,750	1	36,750	24,877	,000
	Level 2 vs. Level 3	48,000	1	48,000	17,600	,001
Error(SITUATIE)	Level 1 vs. Level 2	16,250	11	1,477		
	Level 2 vs. Level 3	30,000	11	2,727		

Figura 4.17. *Output SPSS 11.0 cu testul contrastelor standardizate „Repeated"*
SPSS este o marcă înregistrată.

În cazul primei ipoteze se obține un F (1, 11) = 24,87, p < .01. Astfel, aplicarea unor categorii identice pentru stimuli similari conduce la accentuarea distanței percepute față de stimulii diferiți. Media distanței dintre stimuli în cazul aplicării etichetei a fost 18,75, în timp ce media distanței dintre stimuli în grupul de control a fost 17,00. Similar, în cel de-al doilea contrast, F (1, 11) = 17,60, p < .01. Acest rezultat arată că aplicarea unor categorii identice pentru stimuli diferiți conduce la reducerea distanței percepute dintre stimulii diferiți (media pentru grupul de control a fost 17,00, iar cea pentru ultimul grup experimental a fost 15,00). Se constată că pragul scăzut de semnificație a celor două contraste (.000, respectiv .001) elimină problema unei valori inflaționiste a lui p, peste valoarea de .05, datorată faptului că acest contrast „repetat" nu este ortogonal. În aceste condiții, ambele ipoteze de lucru au rămas în picioare, sprijinind premisa generală a cercetării de față, conform căreia gruparea pe categorii a stimulilor afectează modul de percepție a acestora.

Testele post hoc pot fi utilizate şi în cazul ANOVA cu măsurători repetate.

Pairwise Comparisons

Measure: MEASURE_1

(I) SITUATIE	(J) SITUATIE	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	1,750*	,351	,001	,761	2,739
	3	3,750*	,463	,000	2,445	5,055
2	1	-1,750*	,351	,001	-2,739	-,761
	3	2,000*	,477	,004	,656	3,344
3	1	-3,750*	,463	,000	-5,055	-2,445
	2	-2,000*	,477	,004	-3,344	-,656

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Figura 4.18. *Output SPSS 11.0 cu rezultatele comparațiilor post hoc*
SPSS este o marcă înregistrată.

Totuşi, criteriul de selecție a acestor proceduri se schimbă semnificativ, deoarece specificul design-ului cu măsurători repetate şi neîndeplinirea frecventă a condiției de sfericitate afectează nivelul erorii de tip I. Detalii despre modul de comportare a procedurilor post hoc sunt furnizate în Field (2000), Kinnear şi Gray (2000) şi Weinfurt (2000). Concluzia principală este că cea mai bună opțiune de a testa diferențele dintre grupe este metoda Bonferroni, pentru orice situație existentă. În exemplul dat, rezultatele confirmă diferențele obținute prin contraste.

4.3.1. Mărimea efectelor în ANOVA cu măsurători repetate

În ceea ce priveşte calcularea mărimii efectului, aceasta se realizează într-un mod similar cu cel prezentat în cazul ANOVA simplă şi ANOVA factorială. Spre exemplu, vom calcula mărimea efectului corespunzătoare aprecierii distanței dintre stimuli în situația etichetării comune a stimulilor similari în comparație cu etichetarea comună a stimulilor diferiți, deci cazul grupelor 1 şi 3. Luând în considerare c• F contrast este 65,59 $(3,75/0,463)^2$, iar F (2, 22) este 37,43, vom obține, pe baza formulei 4.3, un r de 0,82 şi în consecință, un r^2 este 0,67. Acești indicatori arată un efect extrem de puternic produs de tipul de etichetare a stimulilor în aprecierea distanțelor dintre ei. Rezultatul are o importanță deosebită în ameliorarea unor situații practice. Astfel, ar trebui descurajată etichetarea diferită a romilor de populația majoritară, pentru a evita accentuarea diferențelor percepute dintre cele două categorii. Similar, utilizarea aceloraşi etichete în descrierea sau numirea persoanelor aparținând celor două etnii, poate determina o reducere importantă a diferențelor percepute între cele două populații.

4.3.2. Condiții necesare pentru aplicarea ANOVA cu măsurători repetate

Primele două condiții din tehnicile ANOVA simplă sau ANOVA factorială se păstrează, deşi neîndeplinirea lor nu afectează foarte mult rezultatele obținute.

În schimb, nu este suficientă îndeplinirea condiției de omogenitate a dispersiilor dintre grupe pentru a aplica ANOVA cu măsurători repetate. Condiția necesară este denumită condiția de sfericitate. Aceasta implică premisa unei relații similare între fiecare pereche de condiții experimentale, ea fiind o condiție mai generală a simetriei complexe. Aceasta din urmă este îndeplinită, dacă dispersia este egală în toate situațiile experimentale (omogenitatea dispersiei) şi dacă covarianța, implicit corelația dintre fiecare condiție experimentală este egală, adică există o dependență similară între toate condițiile experimentale. În practică se observă că este foarte dificil de îndeplinit dubla condiție, majoritatea design-urilor ANOVA cu măsurători repetate cu mai mult de două grupe încălcând această condiție. Să presupunem un design cu patru măsurători repetate. A respecta condiția de simetrie complexă înseamnă că toate cele patru grupe au dispersii

relativ egale şi că există coeficienţi de corelaţie relativ egali între grupele 1 şi 2, 1 şi 3; 1 şi 4; 2 şi 3; 2 şi 4; 3 şi 4. Uneori, aceste condiţii sunt parţial îndeplinite. De exemplu, există o corelaţie relativ egală între rezultatele din grupele 1 şi 2, respectiv 1 şi 3, dar nu şi între 2 şi 3 etc. Această situaţie este numită sfericitate locală.

Din punct de vedere statistic, sfericitatea afectează valorile lui F, crescând şansele de a comite eroarea de tip I, adică de a respinge ipoteza nulă când nu există un efect real. Cum ne dăm seama dacă condiţia de sfericitate este îndeplinită? SPSS-ul calculează testul W al lui Mauchly. Dacă rezultatul acestuia este semnificativ statistic, condiţia de sfericitate este încălcată. Dacă rezultatul nu este semnificativ statistic (adică p este mai mare de .05), putem considera că este respectată condiţia. Totuşi, acest test este foarte sensibil la numărul de subiecţi şi la forma distribuţiei datelor, de aceea este necesară foarte multă precauţie în interpretarea sa (O'Brien şi Kaiser, 1985).

În primul caz, în care condiţia de sfericitate nu este îndeplinită, va trebui să adoptăm una dintre cele două alternative: utilizarea tehnicii MANOVA (analiza dispersională multivariată), sau ajustarea valorilor lui F din ANOVA cu măsurători repetate. Prima nu face obiectul analizei de faţă, în schimb ne vom opri să detaliem cea de a doua posibilitate.

Pentru a înţelege aceste corecţii vom porni de la premisa că îndeplinirea în totalitate a sfericităţii poate fi notată epsilon egal cu 1. De asemenea, cea mai gravă violare a acestei condiţii este egală cu 1/(k-1). În cazul a trei grupe experimentale (k = 3), valoarea minimă a lui epsilon este 0,50. De obicei, un epsilon între .90 şi 1 nu ridică nici un fel de probleme. Cu cât valoarea lui epsilon este mai mică, cu atât sunt probleme mai mari legate de neîndeplinirea condiţiei de sfericitate. Ajustarea lui F se face folosind fie corecţia Greenhouse-Geisser, fie cea concepută de Huynh şi Feldt. Prima modalitate este conservatoare (subestimează valoarea lui epsilon), iar cea de-a doua, liberală (supraestimează valoarea lui epsilon). Specialiştii au preferinţe diferite de a alege una dintre cele două modalităţi de ajustare a lui F. Everitt (1995) consideră că ar trebui realizată o medie între cele două ajustări. Girden (apud Field, 2000) recomandă utilizarea corecţiei Huynh-Feldt, dacă epsilon este cuprins între .75 şi .90, respectiv a corecţiei Greenhouse-Geisser pentru valori epsilon mai mici.

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SITUATIE	,879	1,291	2	,524	,892	1,000	,500

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Figura 4.19. *Output SPSS 11.0 cu rezultatele referitoare la situaţia sfericităţii obţinute*
SPSS este o marcă înregistrată.

Rezultatele din figura 4.19. indică un caz fericit, dar, din păcate rar, de îndeplinire a condiţiei de sfericitate. Valoarea lui epsilon este cuprinsă undeva în intervalul .89 şi 1.00, adică are o valoare mai mare de .90. Acest indiciu este însoţit de un test Mauchly nesemnificativ statistic. În consecinţă, condiţia de sfericitate a fost îndeplinită, nefiind necesare nici un fel de ajustări suplimentare în interpretarea lui F.

Vom folosi aceeaşi situaţie de cercetare pe o altă bază de date imaginare pentru a demonstra modul de lucru în cazul în care condiţia de sfericitate nu este îndeplinită.

Descriptive Statistics

	Mean	Std. Deviation	N
efect de accentuare	18,75	2,491	12
grup de control	18,50	2,236	12
efect de reducere	15,83	4,282	12

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SITUATIE	,341	10,764	2	,005	,603	,636	,500

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
SITUATIE	Sphericity Assumed	62,722	2	31,361	5,478	,012	,332
	Greenhouse-Geisser	62,722	1,205	52,033	5,478	,030	,332
	Huynh-Feldt	62,722	1,273	49,285	5,478	,028	,332
	Lower-bound	62,722	1,000	62,722	5,478	,039	,332
Error(SITUATIE)	Sphericity Assumed	125,944	22	5,725			
	Greenhouse-Geisser	125,944	13,260	9,498			
	Huynh-Feldt	125,944	13,999	8,997			
	Lower-bound	125,944	11,000	11,449			

Figura 4.20. *Valorile descriptive, testarea condiţiei de sfericitate şi tabelul sumar ANOVA cu măsurători repetate pentru aceeaşi problemă, utilizând o altă bază de date*
SPSS este o marcă înregistrată.

În această situaţie se observă că atât testul Mauchly, cât şi valoarea lui epsilon indică o încălcare a condiţiei de sfericitate. Acest aspect poate fi observat empiric prin faptul că abaterea standard a persoanelor aflate în a treia situaţie experimentală este aproape dublă faţă de abaterea standard a celorlalte două grupe. Or abaterea standard nu este altceva decât rădăcina pătrată a dispersiei.

Din al doilea tabel prezentat în figura 4.20. observăm că valoarea lui epsilon este undeva între .603 şi .636, astfel încât sfericitatea este grav încălcată, de aceea vom alege corecția Greenhouse-Geisser a lui F. Ca urmare, în cel de-al treilea tabel al figurii 4.20., nu vom citi prima valoare, adecvată respectării condiției de sfericitate [F(2, 22) = 5,47 la un p de .012], ci vom utiliza valoarea lui F din dreptul corecției Greenhose-Geisser [F(1.20, 13.26) = 5,47, la un p de .03]. În această situație ambele valori F au fost semnificative statistic, însă se observă că după aplicarea corecției pragul lui F a crescut. Acest lucru poate fi important în alte situații în care valorile corectate ale lui F nu sunt semnificative statistic, în timp ce valorile originale ale lui F au fost semnificative, schimbând astfel interpretarea rezultatelor.

Un ultim aspect tehnic, ce merită a fi menționat pe scurt în cadrul analizei dispersionale, ține de modul de calculare a sumei pătratelor (SS). Dacă toate rezultatele sunt complete, vom păstra tipul III de calcul, acesta fiind opțiunea standard în SPSS. Dacă vom avea date lipsă, caz întâlnit mai ales în situația măsurătorilor repetate, vom utiliza tipul IV de calculare a sumei pătratelor.

4.4. Alte modalități ANOVA

Cele trei modalități prezentate mai sus reprezintă cele mai frecvente situații care necesită analiza dispersională. Totuşi, în tabelul 4.8. sunt prezentate alte câteva tehnici ANOVA. Criteriul de diferențiere a lor este dat de independența evaluărilor (grupele formate din aceiaşi subiecți sau din subiecți diferiți) şi de numărul de variabile independente (unul sau mai mulți factori).

Tabelul 4.8. Tipuri de ANOVA în funcție de designul cercetării

	Subiecți diferiți	Aceiaşi subiecți
1 variabilă independentă	ANOVA simplă (ANOVA unifactorială)	ANOVA cu măsurători repetate
2 sau mai multe variabile independente	ANOVA factorială	ANOVA factorială cu măsurători repetate
	Design mixt (unul sau mai mulți factori variați în grupe de subiecți diferiți şi unul sau mai mulți factori măsurați pe aceiaşi subiecți)	

Un design mixt este oferit de Weinfurt (2000). Un cercetător doreşte să observe eficiența terapiei cognitiv-comportamentale în reducerea simptomelor depresive la pacienții suferinzi depresie severă. După o evaluare inițială cu „Inventarul de Depresie Beck", subiecții au fost împărțiți aleatoriu în două grupe: de control şi experimentală. Cei din grupa experimentală au urmat o terapie cognitiv-comportamentală pe o perioadă de trei luni. Inventarul Beck a mai fost administrat

tuturor pacienților la trei luni, la şase luni şi la douăsprezece luni de la prima evaluare. Întrebarea care se pune este dacă pacienții care au urmat terapia cognitiv-comportamentală şi-au redus semnificativ simptomele de depresie pe parcursul acestei perioade de timp.

4.5. Exemplu de ANOVA factorială: cazul euristicii de accesibilitate

În luarea deciziilor sunt utilizate o serie de strategii cognitive denumite euristici. Între acestea este vorba şi de euristica de accesibilitate care constă în tendința oamenilor de a acorda o probabilitate mai ridicată alternativei sau evenimentului care este mai uşor de amintit. Imaginați-vă că sunteți rugat să oferiți cincisprezece exemple în care v-ați purtat agresiv. Probabil că veți oferi rapid şapte-opt exemple după care vă va fi din ce în ce mai greu să vă amintiți celelalte situații până la cincisprezece. Datorită acestui fapt, probabil, nu vă veți evalua ca o persoană foarte agresivă. Imaginați-vă că sunteți rugat să oferiți cinci exemple în care v-ați comportat agresiv. Le veți oferi rapid şi veți concluziona, probabil, că sunteți o persoană foarte agresivă. În aceste condiții se întâmplă un paradox, persoanele din prima situație se vor autoevalua ca fiind mai puțin agresive decât persoanele din al doilea grup, deşi primele au oferit de trei ori mai multe exemple de agresivitate. Acest lucru se produce datorită experimentării dificultății de a rememora toate exemplele tip din prima situație în comparație cu uşurința accesării lor în a doua situație, în care trebuia prezentate doar cinci exemple. În aceste condiții, vom concepe un experiment similar cu cel realizat de Schwartz (apud Kunda, 1999), dar uşor modificat prin introducerea unei trepte suplimentare.

Unii subiecți incluşi în experiment au fost rugați să ofere cinci exemple de situații în care s-au comportat asertiv. Alți subiecți au fost rugați să ofere zece exemple, iar o a treia grupă de participanți la studiu a fost rugată să ofere cincisprezece exemple de situații în care s-au manifestat asertiv. Prin urmare, prima variabilă independentă manipulată de cercetător este numărul de exemple amintite, cu trei trepte de lucru: cinci, zece, respectiv cincisprezece situații de comportament asertiv.

Alte trei grupe de participanți au fost formate schimbând obiectul reamintirii. Astfel, subiecții din a patra grupă au fost rugați să-şi amintească cinci exemple de situații în care s-au comportat non-asertiv, cei din grupa cinci au fost solicitați să-şi amintească zece asemenea situații, în timp ce persoanelor din grupa şase li s-a cerut să prezinte cincisprezece situații în care s-au comportat non-asertiv. Aşadar, cea de-a doua variabilă independentă a constat în obiectul evaluărilor situațiilor stimul, ea având două trepte: asertiv şi non-asertiv. Avem de a face în acest caz cu un design factorial de tip 3 x 2, cu şase celule experimentale, fiecare fiind alcătuită din subiecți diferiți. Variabila dependentă a constat într-o singură întrebare de autoevaluare a gradului de asertivitate a subiectului. Întrebarea este pusă pe o scală Likert cu zece puncte, situate de la 1 (deloc asertiv) la 10 (foarte asertiv).

Rămâne de analizat în ce măsură se confirmă ipoteza conform căreia: persoanele care întâmpină dificultăți în reamintirea situațiilor cerute consideră că au un nivel mai scăzut de dezvoltare a acelei trăsături datorită accesibilității scăzute a unor exemple (euristica de accesibilitate).

Orice analiză dispersională necesită parcurgerea următorilor paşi:

- ✓ selectarea modalității de analiză ulterioară testului F: metode a priori sau comparații post hoc;
- ✓ testarea condițiilor aferente tipului de ANOVA efectuată;
- ✓ interpretarea rezultatelor obținute.

Minimum de informații care trebuie trecute în secțiunea de analiză cantitativă a datelor atinge următoarele puncte:

- ✓ precizarea într-un paragraf a tipului de analiză întreprinsă (ANOVA factorială, ANOVA simplă etc.) şi a justificării pentru alegerea făcută, incluzând aici o discuție scurtă asupra condițiilor;
- ✓ prezentarea unui tabel sumar ANOVA, cât şi a unui tabel cu mediile şi dispersiile fiecărei celule experimentale - vezi modelul de mai jos inspirat din Nicol şi Pexman (1999);
- ✓ în cazul unui număr mare de teste F efectuate se poate renunța la prezentarea tabelului sumar ANOVA, valorile F obținute fiind trecute în text;
- ✓ reprezentarea grafică după caz, fie a relațiilor dintre factori (în cazul unei interacțiuni semnificative), fie a efectelor principale care sunt semnificative;
- ✓ discutarea rezultatelor de la testele post hoc sau a contrastelor, în funcție de caz;
- ✓ interpretarea sumară a rezultatelor găsite.

Pentru exemplul dat, secțiunea de analiză a datelor poate fi formulată în modul următor:

Media şi abaterea standard a asertivității auto-percepute în funcție de stimulul dat şi numărul de exemple cerute

	Dovezi de asertivitate		Dovezi de non-asertivitate	
Număr exemple	M	AS	M	AS
Cinci	6,40	0,96	4,30	0,94
Zece	5,30	0,94	5,20	0,78
Cincisprezece	4,50	1,17	6,00	1,05

Tabelul sumar ANOVA

Sursa dispersiei	SS	df	MS	F
Stimulul dat	0,13	2	0,065	0,06
Număr exemple	0,81	1	0,810	0,83
Stimul dat X Număr exemple	32,53	2	16,265	16,66 **
Reziduală	52,70	54	0,976	
Total	86,18	59		

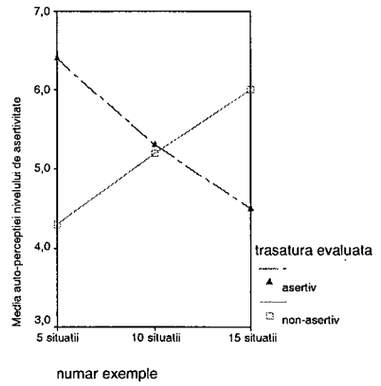
** $p < .01$

Pentru a testa în ce măsură dificultățile de reamintire a unor situații definitorii pentru comportamentul asertiv afectează gradul de auto-percepere a unei persoane ca fiind asertivă, vom apela la calcularea lui F din ANOVA factorial. Am ales această tehnică datorită design-ului factorial de tip 3 x 2, cu tipul de stimul (asertiv vs. non-asertiv) şi numărul de exemple cerute (5, 10 şi 15) ca variabile independente. Toate condițiile specifice aplicării probei au fost îndeplinite, inclusiv omogenitatea dispersiei, unde testul Levene a fost $F (5, 54) = 0,23$, $p > .10$.

Valorile descriptive pentru fiecare situație experimentală sunt prezentate în primul tabel, iar rezultatele obținute în urma aplicării testului ANOVA 3 x 2 sunt trecute în cel de-al doilea.

Din tabelul sumar ANOVA se observă că nici unul dintre efectele principale nu au fost semnificativ statistic. În aceste condiții, orice comparație post hoc sau orice contrast a priori nu-şi mai găsesc rostul. În schimb, efectul determinat de interacțiunea dintre cei doi factori a fost semnificativ statistic, $F (2, 54) = 16,66$, $p < .01$, indicând o interdependență puternică între aceştia Situația specială poate fi observată eficient şi în figura de mai jos.

Din analiza ei putem observa că ipoteza de lucru este susținută. Astfel, persoanele care trebuiau să-şi amintească cinci exemple în care s-au comportat asertiv s-au perceput pe sine ca fiind mai asertive decât persoanele care trebuiau să-şi amintească cincisprezece situații de acest gen. Mai mult, persoanele care trebuiau să-şi amintească cinci exemple în care s-au comportat non-asertiv au urmat un proces similar, dar invers. Ele au obținut un scor mai mic la asertivitate în situația în care trebuiau să-şi amintească cinci stimuli, în comparație cu situația în care trebuiau să-şi amintească cincisprezece asemenea situații, deoarece au întâmpinat mai multe dificultăți în amintirea situațiilor în care s-au comportat non-asertiv. În schimb, nu par a fi diferențe majore între cele două situații-stimul în cazul de mijloc, de reamintire a zece situații stimul. Aceste date sunt sprijinite şi de faptul că patru din cele cinci efecte simple sunt semnificative statistic, singurul efect care n-a îndeplinit acest criteriu fiind tocmai situația referitoare la amintirea a zece stimuli asertivi sau non-asertivi.



Evoluția nivelului de asertivitate auto-percepută în funcție de condițiile experimentale

Rezultatele sugerează ideea că oamenii tind să fie influențați în percepțiile lor de dificultatea de a accesa informațiile cerute (este mai dificil să-ți aduci aminte cincisprezece situații stimul decât cinci) şi mai puțin de numărul de informații aduse. Cu alte cuvinte, chiar dacă ni se cere să precizăm cât mai multe situații în care să ne fi comportat asertiv, dacă vom întâmpina dificultăți în amintirea acelor situații, vom fi tentați să ne caracterizăm în direcție opusă. În cazul de față, vom evalua ca fiind mai degrabă non-asertivi. Astfel, a fost dovedită încă o dată influența euristicii de accesibilitate în modul de evaluare a situațiilor

4.6. Tehnicile ANCOVA

Nu întotdeauna variabilele independente de interes din cadrul unui studiu comparativ sunt de tip nominal (ex. bărbați sau femei) ori ordinal (ex. grup control, grup placebo, grup tratament terapeutic). În multe situații, factorii pot fi măsurați de scale numerice, fie ele de tip interval sau proporții. Inteligența exprimată în valori IQ, trăsăturile de personalitate măsurate prin anumite teste sunt numai câteva cazuri în care datele sunt măsurate pe un interval larg. A lua fiecare valoare numerică exprimată şi a face din ea o treaptă a variabilei independente se dovedeşte a fi o măsură imposibilă de adoptat practic, fie şi numai din motivul necesității unui număr imens de subiecți. Să ne gândim la situația unei testări psihologice, în urma căreia avem 40 de valori numerice diferite. Nu putem realiza patruzeci de grupe de comparație, deoarece acest efort ar fi prea imens şi lipsit de eficiență în comparație cu efortul depus. Dacă factorii măsurați pe o scală numerică continuă sunt importanți în clarificarea relațiilor de influență între VI şi VD trebuie să ținem cont de ei în studiu. Cum putem face acest lucru în situația în care nu ne putem permite şi nici nu are sens să realizăm grupe de subiecți pentru fiecare valoare numerică obținută? Două variante de răspuns, utilizate frecvent de cercetători, vor fi prezentate în cele ce urmează. O a de-a treia modalitate constă în utilizarea regresiei ca metodă de analiză şi va fi discutată în capitolul destinat regresiei liniare.

(1) Prima metodă constă în stabilirea a două sau trei trepte ale variabilei independente pe baza condensării rezultatelor obținute. În cazul realizării a două trepte, subiecții pot fi împărțiți în două grupe, utilizând valoarea medianei ca unitate de demarcație între cele două categorii de subiecți. Prin această metodă, nu va fi pierdut nici un subiect, iar cele două grupe vor fi împărțite în mod egal: grupa cu rezultate inferioare medianei (ex. abilități scăzute) şi grupa cu performanțe superioare medianei (ex. abilități crescute). Totuşi, această împărțire are marele dezavantaj de a distorsiona într-o anumită măsură datele. Gândiți-vă la o situație în care ar fi vorba despre extroversiune, iar valoarea medianei a fost 12,7. În acest caz subiectul care a obținut valoarea 12 a fost pus într-o grupă, iar cel care a obținut valoarea 13 a fost distribuit în cealaltă grupă. Cunoscând faptul că orice instrument de măsurare are anumite erori de măsurare, reflectate în faptul că nu

are o fidelitate perfectă, apare riscul de a grupa greşit un individ care a obținut un rezultat foarte apropiat medianei. În plus, el va fi tratat la fel ca o persoană care a realizat scoruri extreme, pierzându-se foarte mult din nuanțările necesare existente prin măsurarea caracteristicii pe o scală numerică. În aceste condiții, o altă variantă poate fi utilizată pentru gruparea subiecților în trepte ale variabilei independente. De această dată sunt utilizate valorile mediei şi abaterii standard pentru a categorisi subiecții. Varianta elimină multe dintre neajunsurile prezentate anterior, dar are un alt dezavantaj major, şi anume: eliminarea unor subiecți testați din analiza ulterioară a datelor, fapt ce nu ni-l putem permite în multe situații. Totuşi, pentru a exemplifica modalitatea de împărțire a subiecților vom porni de la cazul măsurării gradului de extroversiune, unde media a fost 12, iar abaterea standard 4. Grupa „extroversiune scăzută" ar putea fi alcătuită din subiecții care au obținut rezultate mai mici sau egale cu [media – 0,5 abateri standard], adică din subiecții cu un grad de extroversiune cuprins între 0 şi 10. Grupa „extroversiune ridicată" ar putea fi alcătuită din subiecții care au obținut rezultate egale sau mai mari decât [media + 0,5 abateri standard], adică din persoane care au avut rezultate egale sau mai mari decât 14. Astfel, subiecții cuprinşi în intervalul m±0,5s nu vor fi cuprinşi în nici una dintre grupe şi vor fi eliminați din analiza ulterioară a datelor. În cazul prezentat ar fi fost vorba de toți subiecții care au obținut valorile 11, 12 sau 13 la testul de extroversiune. În cazul în care distribuția rezultatelor ar fi fost normală, ar fi însemnat eliminarea a aproximativ 38% din subiecți. În plus, numărul de subiecți din cele două grupe nou-formate ar fi fost egal doar în condițiile unei distribuții perfect simetrice. Ținând cont de aceste limite descrise, alți cercetători, şi includ aici majoritatea specialiştilor în metodologie, susțin păstrarea, pe cât posibil, a datelor culese inițial şi analizarea lor prin intermediul tehnicilor ANCOVA. Totuşi, în cazurile în care condițiile de utilizare a tehnicii ANCOVA nu sunt îndeplinite, varianta împărțirii subiecților în două sau trei grupe, proces denumit grupare (în engleză „blocking"), rămâne o soluție viabilă, însă designul cercetării nu va mai fi complet randomizat, ci va deveni de tip grupat randomizat (în engleză „randomized block design") (Kirk, 1982).

(2) Analiza de covarianță (ANCOVA) reprezintă o extensie a tehnicilor ANOVA prezentate anterior. În contextul design-urilor de tip ANOVA, introducerea unei covariabile numerice conduce, de cele mai multe ori, la o creştere a puterii de decelare a efectului unei variabile independente asupra uneia dependente. Aceasta se poate face prin controlul statistic al unor diferențe de interes dintre subiecții unei cercetări. Simplu spus, ANCOVA nu este altceva decât o ANOVA în care efectele principale şi interacțiunile dintre factori sunt evaluate după ce valorile variabilei dependente au fost ajustate în funcție de valorile covariabilei. Dacă covariabila ar fi gradul de inteligență, VD ar fi abilitatea de a citi, iar VI ar fi tipul de program educațional oferit, ANCOVA ar evalua efectul acestor programe asupra abilităților de citire în situația ipotetică în care toți subiecții ar avea acelaşi nivel de inteligență. Astfel, în ANCOVA valorile VD sunt mai întâi ajustate în funcție de covariabilă şi apoi, analizate într-un mod similar ANOVA.

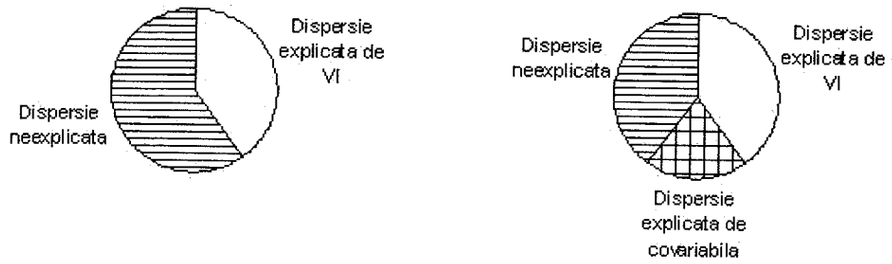


Figura 4.21. *Sursa dispersiei în situația ANOVA (în stânga), respectiv ANCOVA (în dreapta)*

ANCOVA poate fi utilizată atât în sensul de a identifica un efect cât mai „pur" al VI asupra VD, cât şi de a reduce dispersia datorată erorii. Mai precis este vorba de a elimina din calculul F acea parte a dispersiei VD care se datorează unui alt factor, ce inițial nu a fost luat în calcul. Spre exemplu, dacă am măsura efectele diferitelor programe educaționale (ex. Step by Step, Waldorf, tradițional) asupra performanțelor de a citi, ar fi bine să controlăm nivelul de inteligență al copiilor, dat fiind cunoscută influența acestui factor asupra abilității de a citi. Dacă inteligența este măsurată numeric pe o scală de la 50 la 130, ea nu poate fi introdusă ca o altă variabilă independentă, ci va fi tratată ca o covariabilă prin ANCOVA. În urma acestui demers am putea evalua care a fost efectul manipulării VI asupra VD, după ce am elimina influența asupra VD determinată de nivelul intelectual diferit al copiilor incluşi în studiu. Textual, prin ANCOVA am răspunde în acest caz la întrebarea: ținând la nivel constant influența inteligenței asupra abilității de a citi, există diferențe semnificative între cele trei programe de educație în ce priveşte abilitatea de a citi? De asemenea, de multe ori, prin introducerea acestei covariabile s-ar reduce dispersia erorii şi va creşte, astfel, puterea experimentului, înțeleasă ca o capaci-tate de a decela existența sau inexistența unui efect. Să ne reamintim că analiza dispersională nu face altceva decât un raport între dispersia explicată şi dispersia erorii. Astfel, pentru a obține un F semnificativ, prima componentă ar trebui să fie cu un anumit grad mai mare decât cea de-a doua. Prin introducerea covariabilei, o parte a dispersiei erorii (neexplicate) devine explicată. Figura 4.21. prezintă cel mai bine modul de partiționare a dispersiei în ANOVA față de ANCOVA.

De asemenea, ANCOVA este deosebit de răspândită şi în design-urile cvasi-experimentale de tip pretest – postest, implicând o grupă experimentală şi una de control. Un exemplu tipic ar fi situația în care două clase paralele de elevi sunt măsurate inițial în privința gradului de coeziune a grupului din care fac parte. Apoi, una dintre clase este supusă unor tehnici de dinamică de grup menite să îmbunătățească climatul şi unitatea clasei. După finalizarea intervenției psihologice, cele două clase vor fi testate din nou asupra gradului de coeziune a

clasei de elevi. Într-o astfel de situație, randomizarea inițială a elevilor nu este posibilă, cele două grupe fiind constituite pe baza criteriului natural, acela al păstrării locului fiecărui elev în clasa din care face parte. În asemenea situații, diferențele dintre clase pot apărea de la început, înainte de aplicarea intervenției psihologice. Prin utilizarea rezultatelor de la pretest pe post de covariabilă, vom putea răspunde la întrebarea: se dovedeşte eficientă intervenția în condițiile în care cele două grupe ar fi pornit de la condiții egale în ce priveşte gradul de coeziune a clasei de elevi?

Pentru a putea oferi acest răspuns, pe baza relației dintre covariabilă şi VD se obțin valorile ajustate ale mediei grupelor incluse în cercetare pe baza unei ecuații de regresie. De remarcat că tehnicile ANCOVA nu se limitează la o singură covariabilă introdusă. Astfel, cu respectarea anumitor condiții ce vor fi amintite spre finalul acestui capitol, pot fi incluse mai multe covariabile. De asemenea, acestea pot fi introduse în diferite design-uri ANOVA, determinând, într-o manieră similară, situații de ANCOVA simplă, ANCOVA cu măsurători repetate, ANCOVA factorială sau ANCOVA mixtă. În cele ce urmează, ne vom rezuma la un exemplu de ANCOVA simplă, implicând o singură VI, o covariabilă şi o VD.

Situația concretă se bazează pe studiul efectuat de Din şi Calao (2001), modificat însă din punctul de vedere al design-ului cercetării. Cei doi au investigat în ce măsură copiii care au utilizat jocuri educaționale pe calculator au progresat mai repede în ceea ce priveşte dobândirea cititului în comparație cu copiii care nu au beneficiat de această facilitate. Astfel, acest design constă în două grupe de subiecți. Variabila dependentă a constat într-un test de evaluare a progresului în citire. Deoarece acest aspect era dependent şi de un anumit nivel de dezvoltare a nivelului intelectual, variabila inteligență generală, măsurată prin WISC a fost inclusă ca o covariabilă.

Baza de date aferentă acestui exemplu este ancova simpla.sav, iar calea de urmat în SPSS pentru a realiza ANCOVA simplă sau ANCOVA factorială este „Analyze" - „General Linear Model" - „Univariate".

Tabelul 4.9. Tabelul sumar ANCOVA

Sursa dispersiei	SS	df	MS	F
Nivel intelectual	2566,97	1	2566,97	91,28 **
Jocuri calculator	271,88	1	271,88	9,67 **
Reziduală (Eroare)	1040,53	37	28,12	
Total	3879,38	39		

** $p < .01$

Din tabelul 4.9. se observă atât influența semnificativă a nivelului intelectual asupra abilității de a citi, cât şi faptul că cele două grupe de copii incluse în studiu diferă semnificativ în ceea ce priveşte abilitatea lor de a citi, chiar şi după ce a fost controlată influența inteligenței asupra variabilei dependente. De asemenea, se observă că în ANCOVA nu se calculează interacțiunea dintre VI şi

covariabilă. De altfel, una dintre condițiile utilizării ANCOVA presupune tocmai lipsa unei interacțiuni semnificative între cele două.

Din următorul tabel se pot observa uşoarele ajustări ale mediei pentru cele două grupe după menținerea constantă a inteligenței. Cea de-a doua coloană conține valorile reale ale mediilor, obținute în urma aplicării probelor. Cea de-a patra coloană conține, în schimb, valorile celor două medii în situația ipotetică în care nu ar exista diferențe între subiecții din cele două grupe referitoare la nivelul intelectual. Acestea sunt mediile care sunt luate în considerare pentru a constata eventuale diferențe semnificative între populațiile din care au fost extrase grupele în cazul ANCOVA. Se poate constata că, în situația ipotetică în care inteligența copiilor cuprinşi în studiu ar fi constantă, diferența dintre cele două grupe ar scădea uşor, dar ar rămâne semnificativă statistic, în sensul că persoanele care au beneficiat de jocuri educative pe calculator au obținut performanțe superioare celor care nu au avut la dispoziție această facilitate.

Tabelul 4.10. Valorile mediei reale, a celei ajustate şi a abaterii standard reale în cazul abilității de a citi

	Date reale (neajustate)		Date ajustate	
Grupă copii	M	AS	M	AS
Jocuri educative pe calculator	80,45	8,43	79,76	*
Control	73,85	10,90	74,53	*

* SPSS-ul nu oferă valorile abaterii standard pentru datele ajustate

Bineînțeles că în cazul în care VI are mai mult de două trepte vom putea utiliza, atât metoda contrastelor, cât şi cea a comparațiilor post hoc pentru a identifica grupele care diferă între ele. În cazul ANCOVA, procedurile post hoc adecvate sunt Bonferroni şi Sidak, indiferent de numărul de subiecți din grupe sau de omogenitatea dispersiei (Field, 2000). Rămâne de precizat că, în problema dată, pentru a obține atât valorile inițiale ale mediilor, cât şi valorile ajustate ale acestora este necesar să bifăm, în cadrul ferestrei „Options" din SPSS, „descriptive statistics", respectiv „display means for grupa".

4.6.1. Mărimea efectului în tehnicile ANCOVA

Fiind vorba de efecte concrete în plan educativ, vom calcula mărimea efectului pe care-l au jocurile de calculator asupra dezvoltării abilității de a citi a copiilor. Aşa cum am afirmat de mai multe ori pe parcursul acestui capitol, un F semnificativ statistic nu înseamnă neapărat o importanță practică a efectului găsit. Diferențele dintre două grupe pot fi semnificative, dar totuşi minuscule. Astfel, ținând cont că există doar două trepte ale variabilei constituite din persoane care au beneficiat sau nu de jocurile pe calculator, vom utiliza formula 4.2. prezentată în cadrul calculului mărimii efectului în ANOVA simplă. Rezultatul obținut indică un r

de 0,45, iar $r^2 = 0,20$. Dac• ar fi s• utiliz•m indicele d al lui Cohen am obține d = 1,02. Toate aceste valori indică un efect pozitiv important (mare). În consecință, dacă rezultatele studiului sunt confirmate şi în alte ocazii, se demonstrează un fapt esențial: jocurile educative pe calculator au o influență foarte bună asupra dezvoltării abilității de a citi.

Analiza de covarianță, în ciuda avantajelor prezentate mai sus, are o serie de limite şi precauții de care trebuie să ținem cont atunci când abordăm posibilitatea de a aplica această tehnică. Alături de limitele şi condițiile specifice fiecărui tip de design ANOVA pe care le-am prezentat la momentul oportun, există altele suplimentare şi particulare cazului ANCOVA.

4.6.2. Condiții suplimentare aplicării ANCOVA

Mok şi Wheldall (1995) sau Tabachnick şi Fidell (1996) au oferit o serie de argumente menite să atenționeze cercetătorii asupra utilizării, fără precauții, a tehnicii ANCOVA. Neluarea în considerare a condițiilor necesare aplicării analizei de covarianță poate conduce la ajustări eronate ale mediilor şi în consecință, la concluzii eronate ale studiilor. Trei condiții suplimentare par a fi deosebit de importante în această situație.

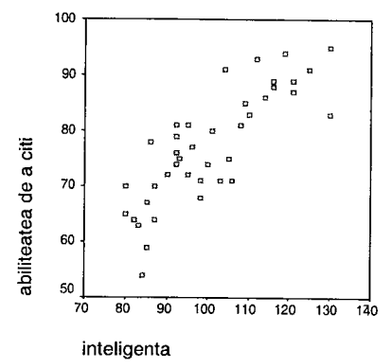


Figura 4.22. *Norul de puncte (scatter-ul) între variabila dependentă şi covariabilă*
SPSS este o marcă înregistrată.

1. Între covariabilă(e) şi variabila dependentă există o corelație liniară. Dacă nu există o corelație sau relația dintre cele două caracteristici nu este liniară, rezultatele ajustării mediilor sunt eronate, deoarece întregul demers ANCOVA se bazează pe regresia liniară. Bineînțeles că metoda cea mai simplă de a verifica această condiție este de a observa norul de puncte rezultat din asocierea celor două variabile şi de a calcula un simplu coeficient de corelație r Bravais-Pearson. Pentru exemplul nostru am obținut o corelație $r(38) = 0,82$, $p < .01$ între nivelul intelectual şi abilitatea de a citi, iar norul de puncte indică o relație liniară între cele două caracteristici.

În cazul utilizării mai multor covariabile, condiția se păstrează pentru fiecare pereche covariabilă – variabilă dependentă în parte. În plus, între covariabile nu trebuie să existe corelații foarte ridicate, datorită fenomenului de multicoliniaritate (vezi secțiunea 6.5.4.).

2. Fidelitatea probei de măsurare a covariabilei trebuie să fie ridicată. O măsurare fără erori este posibilă numai în situații foarte rare, în care variabila este dată de vârsta subiecților, înălțimea lor etc. Cele mai multe teste psihometrice au un nivel de fidelitate care nu atinge niciodată valoarea unu. Conform lui Tabacknick şi Fidell (1996), în cazul utilizării ANCOVA în situații experimentale se ajunge la valori F conservatoare, ceea ce creşte şansa de a comite eroarea de tip II. Acest lucru nu se păstrează şi în cazul situațiilor cvasi şi non-experimentale, în care putem întâlni atât sub-ajustări, cât şi supra-ajustări ale mediilor inițiale. De aceea, autoarele recomandă utilizarea covariabilelor în situații cvasi şi non-experimentale doar dacă fidelitatea măsurării lor atinge cel puțin 0,80. O valoare mai mică introduce prea multă eroare în estimarea mediilor ajustate şi, implicit, în calcularea diferențelor dintre acestea.

3. Efectul covariabilei asupra variabilei dependente este similar pentru oricare treaptă a variabilei independente. Această condiție este cunoscută sub numele de omogenitatea pantei de regresie şi face referire la absența unei interacțiuni semnificative între variabila sau variabilele independente şi covariabilă(e). Unii autori, precum Dugard şi Todman (1995) consideră că ANCOVA este destul de robustă la încălcarea acestei condiții, dacă numărul de subiecți din grupe este relativ egal şi dacă variabila dependentă este distribuită normal. Totuşi, în alte situații, respectarea condiției are o importanță extremă.

Cele două grafice prezentate în continuare exprimă ambele situații: în stânga, de omogenitate a regresiei; în dreapta, de heterogenitate a regresiei (încălcarea condiției). În cazul în care există o omogenitate a regresiei, liniile ce marchează panta de regresie sunt paralele, acest fapt semnificând coeficienți b de

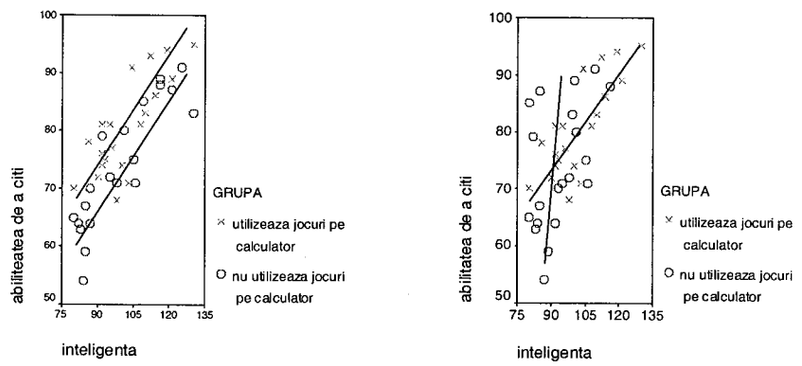


Figura 4.23. *Reprezentarea grafică a situațiilor de omogenitate (stânga) şi eterogenitate (dreapta) a regresiei*

SPSS este o marcă înregistrată.

regresie similari pentru a surprinde relația dintre covariabilă şi variabila dependentă la ambele trepte ale variabilei independente. În situația de heterogenitate a regresiei, pantele nu sunt paralele, prin urmare coeficienții b de regresie vor fi diferiți pentru cele două grupe de copii. În acest caz, nu putem aplica ANCOVA, deoarece ajustarea mediilor se face pe baza unui b general, obținut prin luarea în considerare a tuturor treptelor VI, iar acesta nu este reprezentativ pentru nici una dintre treptele VI.

Deoarece desenul din dreapta, reprezentând cazul eterogenității regresiei este bazat pe date simulate, iar cel din stânga este dat de datele reale ale exemplului oferit, condiția de omogenitate a fost respectată şi ANCOVA a putut fi realizată în exemplul nostru.

O modalitate mai exactă de testare a acestei condiții constă în calcularea unui F de interacțiune între VI şi covariabilă concomitent cu calcularea efectelor principale. Un rezultat semnificativ statistic indică încălcarea condiției de omogenitate a dispersiei, fapt ce împiedică realizarea ANCOVA. O altă posibilitate de testare presupune rularea unei regresii ierarhice în doi paşi. Mai întâi sunt introduse doar covariabila şi variabila independentă (ca variabilă dummy) pe post de predictori; apoi atât covariabila şi variabila independentă, cât şi o variabilă nou formată din produsul celor două, după transformarea prealabilă a variabilei independente în variabilă de tip dummy (vezi secțiunile 6.3. şi 6.7.2.). Dacă „F step", ce măsoară diferența apărută prin introducerea variabilei-produs este semnificativ statistic, există o interacțiune între covariabilă şi factor, deci este încălcată condiția de omogenitate a dispersiei.

4.7. Tipuri de factori în design-urile de tip ANOVA sau ANCOVA

În încheierea acestui capitol, vom face câteva scurte precizări de ordin metodologic menite să faciliteze interpretarea corectă a rezultatelor obținute din ANOVA. Aceasta nu se poate face fără o înțelegere adecvată a tipurilor de factori posibili. Vom folosi trei criterii de clasificare a variabilelor independente: scala de măsurare utilizată; gradul de generalizare a rezultatelor şi gradul de manipulare a factorilor.

(1) După primul criteriu, putem distinge între **factori măsurați nominal, ordinal sau numeric** (incluzând aici atât scala de tip interval, cât şi cea de proporții). Categoriile nominale ale unui factor pot constitui treptele variabilei independente, fiind astfel adecvate tuturor tipurilor de analiză dispersională, cu excepția utilizării contrastelor polinomiale, care necesită o ordonare a treptelor factorilor. Categoriile ordinale ale unui factor pot fi supuse tuturor tipurilor de ANOVA, inclusiv contrastelor polinomiale. În sfârşit, variabilele numerice nu pot fi folosite ca trepte ale variabilei independente în ANOVA decât prin grupări grosiere ale datelor în câteva categorii. Totuşi, prin anumite ajustări, factorii relevanți

măsurați prin scale numerice pot fi utilizați în analize de covarianță - ANCOVA (Pedhazur şi Schmelkin, 1991).

(2) Un al doilea criteriu de diferențiere a factorilor urmăreşte gradul de manipulare a variabilelor şi împarte factorii în: **variabile manipulabile** şi **variabile etichetă** (Kirk, 1982; Radu şi colab., 1993). Astfel, unele variabile, precum nivelul consumului de alcool (ex. 50 mg; 150 mg, 250 mg) pot fi manipulate şi alese de către experimentator. Alte variabile, cum sunt sexul persoanei, vârsta etc. sunt date de caracteristicile subiecților şi nu pot fi manipulate de cercetător. În această situație vorbim despre variabile etichetă. Importanța distincției dintre cele două aspecte ține de subtilitatea interpretării. Astfel, dacă obținem diferențe semnificative statistic între treptele variabilei manipulate, vom trage concluzii directe (ex. un consum de alcool ridicat duce la scăderea semnificativă a vitezei de reacție la stimulii vizuali). În schimb, obținerea unei diferențe semnificative între bărbați şi femei, cu privire la viteza lor de reacție, nu înseamnă prin sine că eticheta de „a fi bărbat" sau „a fi femeie" a condus la această situație. Ceea ce a cauzat diferența dintre sexe ține de anumite aspecte diferențiale care sunt incluse în aceste etichete. Cei orientați spre ştiințele medicale poate vor găsi cantități diferite de anumiți neurotransmițători sinaptici. Unii care vin din sfera antropologiei, vor afirma poate, că diferențele sunt datorate unui stil cultural diferit, bărbații, prin natura lor de vânători, fiind nevoiți să reacționeze mai rapid. Cu alte cuvinte, atunci când avem de a face cu variabile manipulate, cauza este manifestă şi poate fi detectată direct. În schimb, în situația unor variabile etichetă, cauza reală nu este detectabilă, ea fiind ascunsă în spatele multitudinii diferențelor dintre categoriile variabilei etichetă (ex. diferențele dintre bărbați şi femei).

(3) În sfârşit, după cel de-al treilea criteriu, gradul de generalizare a rezultatelor, factorii se împart în **factori ficşi** şi **factori aleatori**. De departe, cel mai întâlnit tip de factori sunt cei ficşi, în care treptele variabilei independente sunt tratate ca măsuri de sine stătătoare. În schimb, factorii aleatori au stabilite trepte care nu sunt identificate ca fiind fixe, ci mai degrabă sunt considerate ca valori reprezentative pentru variabila independentă în cauză (Smith, 2000). Există două criterii care ne ajută să distingem între factorii ficşi şi cei aleatori. Pentru ca un factor să fie considerat aleatoriu trebuie să îndeplinească două condiții de bază (Jackson şi Brashers, 1994): treptele alese să fie arbitrar alese dintr-un eşantion de condiții care se pot interschimba, iar interpretarea rezultatelor să nu prezinte interes la nivelul fiecărei trepte selectate. Pentru a clarifica aceste aspecte, vom apela la următoarele două exemple:

Exemplul 1:

> Un cercetător este interesat să stabilească care dintre următoarele două metode de terapie (cognitiv-comportamentală vs. analiza tranzacțională) este mai eficientă. Din acest punct de vedere, tipul de terapie este un factor fix, deoarece ambele trepte ale tipului de terapie au substanță în sine, cercetătorul fiind interesat în mod special de diferențele dintre cele două tipuri de terapie, deşi există şi alte modalități posibile de intervenție clinică.

Exemplul 2:

> Un cercetător este interesat să stabilească în ce măsură consumul de alcool conduce la creşterea erorilor de execuție. Cercetătorul distribuie subiecții în mod aleatoriu în trei grupe experimentale (50ml, 100 ml şi 200 ml țuică) şi o grupă de control. În această situație, consumul de alcool poate fi tratat ca o variabilă aleatorie deoarece treptele respective pot fi considerate ca valori extrase dintr-un eşantion mai larg. La fel de bine, treptele variabilei independente puteau fi stabilite la 50 ml, 150 ml sau 250 ml. De interes este evoluția erorilor de execuție în urma consumului de alcool, treptele acestuia (50ml, 100 ml şi 200 ml) fiind mai puțin importante, ele fiind alese arbitrar.

Primul exemplu poate fi ulterior complicat prin introducerea unei noi variabile explicative, cum ar fi obiectul intervenției. Să presupunem că factorul respectiv a fost operaționalizat prin două trepte (modalități): persoane care suferă de depresie, respectiv anxietate. Acest factor poate fi conceptualizat atât ca factor fix, cât şi ca factor aleatoriu. El este factor fix dacă cercetătorul este interesat în mod concret de răspunsul la următoarea întrebare: *Terapia cognitiv-comportamentală este mai eficientă în ameliorarea simptomelor pacienților depresivi şi anxioşi decât terapia bazată pe analiză tranzacțională?* Factorul trebuie tratat ca aleatoriu, dacă întrebarea are un caracter general: *Terapia cognitiv-comportamentală este mai eficientă în ameliorarea simptomelor decât terapia bazată pe analiză tranzacțională?*

În acest ultim caz, depresia sau anxietatea nu constituie un deosebit interes, acestea putând fi considerate ca aspecte extrase dintr-un eşantion larg de tulburări psihice precum: tulburările de conversie, cele obsesiv-compulsive etc. Prin urmare, considerarea unei variabile independente ca factor aleatoriu permite o generalizare mai largă a rezultatelor obținute.

Din nefericire, acest avantaj este compensat de riscuri mai mari de a comite eroarea de tip I, deoarece factorii aleatori au o putere statistică de decelare a efectelor existente mai mică decât aceea a factorilor ficşi. În consecință, în cazul unui set de date identice, există o probabilitate mai mare de a sprijini ipoteza cercetării, dacă vom trata factorii ca fiind ficşi.

BIBLIOGRAFIE

Clocotici, V., Stan, A. (2000). *Statistică aplicată în psihologie*. Iaşi: Editura Polirom.

Din, F.S., Calao, J. (2001). The effects of playing educational video games on kindergarten achievement. *Child Study Journal*, 31(2), 95-102.

Dugard, P., Todman, J. (1995). Analysis of pre-test post-test control group designs in educational research. *Educational Psychology*, 15(2), 181-198.

Everitt, B.S. (1995). The analysis of repeated measures: a practical review with exemples. *The Statistician*, 44(1), 113-135.

Eysenck, M.W., Keane, M.T. (2000). *Cognitive psychology. A student's handbook* (4th ed.). Hove, UK: Psychology Press Ltd.

Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage Publications.

Hinkle, D.E., Wiersma, W., Jurs, S.G. (1994). *Applied statistics for the behavioral sciences* (3rd.). Boston, USA: Houghton Mifflin Company.

Jackson, S., Brashers, D.E. (1994). *Random factors in ANOVA*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-098. THousand Oaks, CA: Sage.

Keppel, G. (1991). *Design and analysis: a researcher's handbook* (3rd ed.). Englewood Cliffs, USA: Prentice-Hall Inc.

Kinnear, P.R., Gray, C.D. (2000). *SPSS for Windows made simple. Release 10.0*. Hove, UK: Psychology Press Ltd.

Kirk, R.E. (1982). *Procedures for the behavioral sciences* (2nd ed.). Belmont, USA: Brooks / Cole Publishing.

Kunda, Z. (1999). *Social Cognition. Making sense of people*. London, UK: Massachusetts Institute of Technology.

Lyons, E. (1998). Social psychology 1. În M. Eysenck (ed.). *Psychology: an integrated approach*. Harlow, UK: Prentice Hall Inc.

Mok, M., Wheldall, K. (1995). Some reservations about the use of the analysis of covariance in educational research: a response to Dugard and Todman. *Educational Psychology*, 15(2), 199-202.

Nicol, A.A.M., Pexman, P.M. (1999). *Presenting your findings. A practical guide for creating tables*. Washington DC, USA: American Psychological Association.

O'Brien, R.G., Kaiser, M.K. (1985). MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychological Bulletin*, 97(2), 316-333.

Pedhazur, E.J., Schmelkin, L. (1991). *Measurement, design, and analysis: an integrated approach*. Hillsdale, USA: Lawrence Erlbaum Associates, Inc.

Rosenthal, R., Rosnow, R.L., Rubin, D.B. (2000a). *Contrasts and effect sizes in behavioral research. A correlational approach*. Cambridge, UK: Cambridge University Press.

Rosenthal, R., Rosnow, R.L., Rubin, D.B. (2000b). Contrasts and correlations in effect-size estimation. *Psychological Science*, 11(6), 446-453.

Radu, I., Miclea, M., Albu, M., Nemeş, S., Moldovan, O., Szamoskozi, Ş. (1993). *Metodologie psihologică şi analiza datelor*. Cluj-Napoca: Editura Sincron.

Sava, F. (2002). *Pagina de statistică socială*. Găsit la: http://statisticasociala.tripod.com

Smith, E. (2000). Research design. În H.T. Reiss şi C.M. Judd (eds.). *Handbook of research methods in social and personality psychology*. Cambridge, UK: Cambridge University Press.

Tabachnick, B.G., Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York, USA: Harper Collins College Publishers.

Weinfurt, K.P. (2000). Repeated Measures Analyses: ANOVA, MANOVA, and HLM. În L. Grimm şi P.R. Yarnold (eds.). *Reading and understanding more multivariate statistics* (p.317-362). Washington DC, USA: American Psychological Association.

CAPITOLUL 5

ANALIZA FACTORIALĂ

Sub acest termen generic pot fi incluse teoria şi procedeele corespondente menite să diferenţieze anumite dimensiuni latente pe baza analizei corelaţiilor sau covarianţelor dintre mai multe variabile manifeste. Probabil că nici o metodă statistică nu a fost mai strâns legată de aria de interes a psihologiei. De altfel, Cattell era de părere că analiza factorială "este un instrument care poate fi aplicat studiului comportamentului şi ... poate duce la rezultate ale căror obiectivitate şi validitate se ridică la nivelul celor obţinute în ştiinţele naturale" (Stills,1989, p.114). Această încredere imensă acordată analizei factoriale a dus la aplicaţii pe scară largă în psihologie. Să ne amintim, doar, de teoriile personalităţii dezvoltate de Eysenck, Cattell, sau Costa şi McCrae pe baza analizei factoriale sau de întreaga istorie a studiilor din domeniul inteligenţei (ex. factorul g, expresie a inteligenţei generale; cele două componente verbală şi neverbală, ca forme ale inteligenţei). Merită să subliniem faptul că istoricul analizei factoriale în România a început în perioada interbelică, promotor fiind şcoala clujeană de psihologie. De altfel, Nicolae Mărgineanu a scris prima carte de profil în 1938, denumită "Analiza factorilor psihici".

Din păcate, numărul mare de studii efectuate, precum şi diversitatea procedeelor utilizate au condus la o serie de concepţii greşite privitoare la analiza factorială. Alteori, aplicarea unor reţete rigide în extragerea şi analiza factorilor au condus la rezultate eronate. Pedhazur şi Schmelkin (1991) sintetizează foarte bine această imagine, asemănând analiza factorială cu o pădure frumoasă, dar în care te poţi pierde.

Din aceste considerente, este necesară o scurtă introducere a teoriei analizei factoriale pentru a înţelege logica acesteia. Întâi, se cuvine să distingem clar existenţa a două forme fundamentale de analiză factorială, care vor fi tratate pe rând pe parcursul acestui capitol: analiza factorială exploratorie şi analiza factorială confirmatorie. Pe scurt, **analiza factorială exploratorie** (în engleză „exploratory factor analysis" sau EFA) urmăreşte stabilirea unui număr de factori comuni pentru un set dat de variabile şi identificarea naturii psihologice a acestora. În schimb, **analiza factorială confirmatorie** (în engleză „confirmatory factor analysis" sau CFA) testează ipotezele legate de gradul de adecvare a soluţiilor obţinute în urma analizei exploratorii.

Iată două exemple clarificatoare. Un cercetător este interesat de a identifica care sunt dimensiunile latente ale unei scale de 28 de itemi ce vizează afectivitatea negativă. În urma analizei factoriale exploratorii sunt identificaţi patru factori: pesimism / optimism; cinism / încredere în ceilalţi; insatisfacţie / satisfacţie

față de sine; respectiv neliniște / calm. Un alt studiu evidențiază numai trei factori: nemulțumire; cinism, respectiv pesimism. Să presupunem că primul factor extras - gradul de nemulțumire - conține itemii scalelor de neliniște şi insatisfacție din primul studiu. Prin analiza factorială confirmatorie vom putea testa, pe un alt lot de subiecți, care dintre cele două modele pare a fi mai potrivit pentru a caracteriza întreaga scală de afectivitate negativă.

De obicei, analiza factorială exploratorie (EFA) se utilizează în situațiile în care nu există o perspectivă teoretică clară asupra relațiilor dintre anumite fenomene ce necesită analiza factorială. Pe măsură ce sunt acumulate cunoştințe relevante pentru acel domeniu este preferabil să utilizăm analiza factorială confirmatorie: aceasta depăşeşte caracterul exploratoriu în favoarea testării unor ipoteze specifice.

5.1. Analiza factorială exploratorie

Metoda este utilă în atingerea unor obiective generale, mai importante decât construirea unei simple scale. Mă refer la construirea unor teorii, cum sunt cele menționate anterior în domeniul personalității şi în cel al inteligenței. Acest fapt se datorează ideii că analiza factorială nu este doar o tehnică statistică, ci, în primul rând, o teorie asupra realității. De aceea este necesară o scurtă prezentare a teoriei analizei factoriale. În redactarea acesteia, mă voi baza, în special pe introducerea excelentă a lui Tucker şi MacCallum (1997), dar şi pe cele realizate de Wegener şi Fabrigar (2000), respectiv Bryant şi Yarnold (1995).

5.1.1. Teoria analizei factoriale

Într-o cercetare, domeniul şi populația de interes sunt două concepte esențiale care rămân deseori la un nivel implicit de tratare. Prin domeniu se înțelege un ansamblu de fenomene la care se referă studiul respectiv. Acesta poate fi mai larg (ex. personalitatea; abilitățile mintale) sau mai restrâns (ex. neuroticism; hiperkinezie). Populația este dată de totalitatea entităților care pot intra în studiu (ex. toți copiii de vârstă şcolară). De regulă, din rațiuni ce țin de resurse limitate, se folosesc eşantioane de subiecți extrase din populația respectivă. Într-un mod similar, dintr-un anumit domeniu de cercetare vor fi selectate câteva variabile care vor fi incluse în studiu şi măsurate. Variabilele selectate sunt denumite *atribute de suprafață*. Spre exemplu, în domeniul abilităților mintale pot fi utilizate mai multe probe de evaluare: probleme de aritmetică, de vocabular, de percepere a relațiilor spațiale etc. Fiecare dintre aceste probe prezintă un atribut de suprafață. În consecință, în fiecare domeniu de interes pot fi selectate o multitudine de astfel de atribute. Un set de astfel de atribute supuse măsurării formează o baterie. Acelaşi principiu se aplică şi scalelor

de personalitate. Astfel, dacă domeniul de interes este afectivitatea negativă, vom avea mai multe componente (atribute de suprafață) cum ar fi cinismul, pesimismul etc.

În urma testării eşantionului cu bateria aleasă vom sesiza mai multe aspecte. Mai întâi, vom observa că există diferențe individuale între subiecți, rezultatele fiind dispersate în funcție de gradul de dezvoltare a abilităților măsurate. De asemenea, vom observa o anumită asociere (covarianță) între rezultatele obținute de subiecți la două sau mai multe probe. Astfel, o persoană care a obținut un rezultat ridicat la o problemă de înmulțire e foarte probabil să obțină un scor bun şi la o problemă de adunare.

Dat fiind numărul mare de probe, se pot obține „patternuri" de corelații diverse între probele unei baterii. Unele dintre acestea pot fi corelate puternic, iar altele pot fi necorelate. La prima impresie, relațiile stabilite între variabile (atributele de suprafață) pot părea haotice. Analiza factorială vine şi clarifică această situație, identificând o anumită ordine şi structură în date. Acest lucru se face prin găsirea unei explicații succinte privitoare la dispersia şi covarianța observate în atributele de suprafață.

Ideea fundamentală în analiza factorială este postularea existenței *atributelor interne*. Un astfel de atribut face referire la caracteristici neobservabile direct la oameni, dar care determină diferențe între aceştia. Ele se află la baza atributelor de suprafață. De exemplu, am putea vorbi de abilități numerice, abilități verbale etc. ca atribute interne în domeniul abilităților mintale. Acestea nu pot fi măsurate direct, dar pot fi reflectate prin intermediul evaluării atributelor de suprafață. Astfel, am putea afirma că o persoană este superioară alteia în privința abilităților numerice, dacă la o serie de probe ce implică adunare, scădere, înmulțire etc. obține rezultate mai mari decât cealaltă persoană. Aceste atribute interne sunt cunoscute, în general, sub denumirea de variabile latente, iar în cadrul analizei factoriale ele se numesc factori.

Aceste variabile latente pot avea un substrat real sau pot fi doar constructe teoretice care facilitează înțelegerea şi explicarea realității. O analogie poate fi făcută apelând la concepte fizice precum gravitația. Existența acesteia a fost construită ipotetic, pornind de la observații concrete. Toate obiectele au tendința de a cădea pe pământ cu o viteză variabilă în funcție, printre altele, de masa lor. Prin urmare, există o forță nevăzută care acționează şi determină această cădere a obiectelor. Similar, am putea spune că unii elevi tind să rezolve, mai repede şi mai bine decât alții, probleme de matematică, datorită unui nivel mai ridicat al abilităților lor numerice. Altfel spus, *variabilele manifeste sunt influențate sistematic* de anumite variabile latente (atribute interne, factori).

Se impune acum o nouă distincție între două categorii de atribute interne: *factori comuni* şi *factori specifici*. Primii sunt aceia care influențează mai multe atribute de suprafață din baterie (ex. abilitățile numerice afectează rezultatele obținute atât la proba de adunare, cât şi la cea de înmulțire). În schimb, factorii specifici influențează o singură variabilă manifestă. Dinamica dintre cele două

categorii de factori este influențată de modificările survenite în baterie (scală). Astfel, introducerea sau eliminarea unor probe (itemi), poate conduce la transformarea unor factori specifici în factori comuni şi invers. Să presupunem că avem zece probe menite să măsoare doi factori comuni: abilitățile numerice şi relațiile spațiale. Pentru fiecare atribut intern sunt prevăzute cinci probe. Dacă eliminăm toți itemii care fac referire la abilitățile numerice, cu excepția unei singure probleme de adunare, abilitatea numerică devine un factor specific, deoarece se bazează pe un singur atribut manifest. Alături de aceşti factori comuni şi specifici, în evaluarea atributelor de suprafață intervine şi eroarea de măsurare. Aceasta nu ține de atributele interne, constituie mai degrabă, influențe tranzitorii şi nesistematice care afectează rezultatele obținute la atributele de suprafață (ex. oboseală, momentul zilei ales pentru testare etc.). Între fidelitatea unui test şi erorile de măsurare există o legătură inversă. Cu cât un test este mai fidel, erorile de măsurare sunt mai mici şi invers. În sfârşit, erorile de măsurare combinate cu factorii specifici ai fiecărui atribut de suprafață determină *factorul unic*. Astfel, fiecare rezultat manifest poate fi văzut ca fiind o consecință a influenței unuia sau a mai multor factori comuni şi a factorului unic.

Spre exemplu, vom lua ca model răspunsul afirmativ al unei persoane la itemul "Întotdeauna aştept să se întâmple ce este mai rău". din cadrul scalei de pesimism şi vom constata că el este afectat de trei surse de variație. Prima sursă este dată de factorul comun măsurat în scala respectivă, adică gradul de pesimism al persoanei respective. A doua sursă se referă la factorul specific al acestui item, dat de totalitatea aspectelor care îl individualizează față de ceilalți itemi ai scalei. În acest caz, datorită formulării, ar putea fi vorba de efectul de "framing" pus în evidență de Tversy şi Kahneman (Miclea, 1994). Probabil că o formulare diferită, dar similară de genul: "Nu aştept niciodată să mi se întâmple lucruri bune" ar fi condus la rezultate uşor diferite, dată fiind specificitatea formulării. Dacă mai mulți itemi vor fi afectați, prin modul lor de formulare, de aceste efect de „framing", vom obține un alt factor comun determinat de modul în care sunt formulate afirmațiile (ex. în termeni de câştig sau pierdere). În fine, a treia sursă de variație a rezultatelor la itemul respectiv poate fi determinată de erorile de măsurare (ex. fidelitatea scăzută a probei; prezența unor stimuli distractori etc.) Factorul specific alături de erorile de măsurare constituie, împreună, unicitatea dispersiei scorurilor.

Analiza factorială este mai puțin interesată de acest aspect. În schimb, ea ține seama, în mod deosebit, de dispersia şi covarianța rezultatelor datorate factorilor comuni. Această dispersie comună se numeşte comunalitate. Eventualele corelații existente între anumite probe sunt explicate ca fiind determinate de existența unuia sau mai multor factori comuni. În schimb, lipsa corelației sau o corelație scăzută indică existența posibilă a unor factori diferiți. De exemplu, rezultatele diferite obținute la proba de adunare şi la cea de vocabular constituie o dovadă că cele două atribute de suprafață sunt influențate de factori diferiți.

La o scală mai mare, cu cât există mai multe corelații ridicate între atributele măsurate, cu atât ele tind a dovedi existența unui factor comun. La fel,

lipsa de corelații dintre mai multe variabile manifeste constituie un indiciu al absenței unei comunalități între variabilele respective. În plus, într-un anumit domeniu de interes, se porneşte de la premisa că majoritatea atributelor de suprafață sunt influențate de un număr restrâns de factori comuni.

Toată această discuție asupra analizei factoriale porneşte de la ideea unei legături liniare între variabile. Acest aspect merită să fie reținut, deoarece între multe dintre fenomenele realității psihice nu există o relație liniară. În astfel de situații este inadecvată folosirea analizei factoriale, cel puțin în varianta sa clasică.

Aşadar, analiza factorială nu este mai mult decât o teorie despre dinamica relațiilor dintre atributele de suprafață şi cele interne. Ca orice teorie, ea are o mai mare sau mai mică corespondență cu realitatea. Aplicarea ei în mod automat, fără a ține cont de aceste precizări conceptuale, poate conduce la soluții greşite.

Sintetizând, scopul primordial al EFA este de a identifica numărul şi natura factorilor comuni care stau la baza unui set de variabile manifeste. Tehnica este utilizată atât în construirea de scale, cât şi la dezvoltarea unor teorii. Înainte de a aprofunda modul de realizare practică a unei astfel de analize, se impune să facem o ultimă distincție între analiza factorială exploratorie şi analiza componentelor principale (în engleză „principal component analysis" sau PCA).

5.1.2. Analiza componentelor principale vs. analiza factorială

Ambele metode au fost utilizate inadecvat, de nenumărate ori, în principal datorită neînțelegerii diferențelor dintre ele. Confuzia a fost întreținută şi de cazurile numeroase în care soluțiile, obținute pe acelaşi set de variabile, au fost similare. Nu în ultimul rând, unele cărți de profil au eşuat în încercarea de a face o distincție clară între cele două metode.

Prima diferență importantă între analiza factorială şi analiza componentelor principale este determinată de modul diferit de a analiza datele. Să ne reamintim că analiza factorială identifica trei surse ale variației datelor: factorii comuni, cei specifici şi erorile de măsurare. Gruparea ultimelor două surse oferea unicitatea varianței însă interesul central cădea asupra comunalității, ca dispersie explicată de factorii comuni. În schimb, analiza componentelor principale nu recurge la această triplă distincție, iar teoria de la baza ei nu postulează existența unor factori comuni. În PCA, dispersia datelor are la bază două surse: dispersia explicată şi cea datorată erorilor de măsurare (Iacobucci, 1994). În aceste condiții, nu ar fi normal să vorbim despre factori în analiza componentelor principale, devreme ce teoria nu specifică existența acestora. De aceea, se foloseşte termenul substitut de componentă principală. Atât factorii, cât şi componentele principale pot fi denumite generic "eigenvectori". Scopul analizei componentelor principale este extragerea unui număr cât mai mic de eigenvectori, care să explice cât mai mult din totalul varianței.

Dispersia, în cazul celor două metode, poate fi reprezentată grafic în funcție de sursele care compun totalul acesteia:

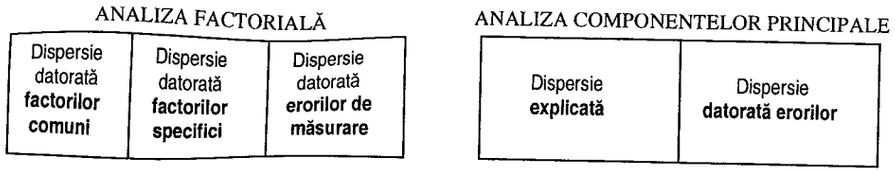


Figura 5.1. *Diferențe conceptuale între analiza factorială şi analiza componentelor principale*

Pornind de la aceste considerente, vom constata a doua diferență esențială între EFA şi PCA. Analiza componentelor principale are ca obiectiv *reducerea datelor* prin identificarea unui număr restrâns de componente principale care explică împreună cât mai mult din totalul dispersiei variabilelor manifeste. Ea face acest lucru pe baza unei combinații liniare între variabilele manifeste. În schimb, analiza factorială exploratorie încearcă să determine un număr restrâns de factori care influențează semnificativ o serie largă de variabile manifeste. Pe scurt, PCA este o tehnică utilă de reducere a datelor, în timp ce EFA este adecvată identificării unor constructe psihologice latente.

O ultimă diferență, subtilă, ține seama de aspectul cauză – efect implicat în cele două analize. În EFA, factorii sunt văzuți ca fiind responsabili de evoluția variabilelor manifeste, asumând rolul unor variabile „cauză". Astfel, pare natural să afirmăm că o persoană obține rezultate ridicate la probe de aritmetică, vocabular, relații spațiale etc. datorită unei inteligențe ridicate. În PCA, componentele principale nu joacă acelaşi rol. Mai degrabă, ele sunt rezultatul combinației liniare a variabilelor manifeste, al cărei scop este de a explica cât mai mult din totalul dispersiei datelor. O situație care necesită PCA este aceea în care dorim să obținem un singur indicator al „stării materiale" pe baza unor informații multiple ce privesc venitul personal al subiectului, venitul financiar al familiei, un index al bunurilor de lux deținute etc. Alte situații în care se aplică PCA sunt cele în care mai multe variabile independente corelează foarte puternic. În acest caz, prin extragerea uneia sau a mai multor componente principale se elimină problema multicoliniarității (vezi secțiunea 6.5.4.). Acestor diferențe între EFA şi PCA se adaugă şi altele, de ordin matematic, implicate în algoritmul de calcul al eigenvectorilor. Asupra lor nu vom insista aici. În plus, trebuie reținut că SPSS-ul oferă doar posibilitatea realizării analizei factoriale. Cei interesați de utilizarea metodei componentelor principale pot apela la alte programe software precum ViSta (Young, 2002) sau LISREL (Jöreskog şi Sörbom, 1989).

În concluzie, deşi numeroase situații dovedesc o similaritate a rezultatelor obținute prin cele două metode, există diferențe semnificative din punct de vedere teoretic. Acestea, la rândul lor, pot duce la rezultate şi interpretări greşite ale soluțiilor obținute. În continuare, vom insista asupra modului de realizare a analizei factoriale. Totuşi, analiza componentelor principale implică parcurgerea aceloraşi etape.

5.1.3. Design-ul şi metodologia analizei factoriale

Obiectivul major al EFA, acela de a identifica factorii care influențează atributele de suprafață, nu poate fi realizat eficient printr-o singură cercetare. Vom avea nevoie de mai multe studii care să conducă treptat la o cunoaştere bună a atributelor interne. Nu se urmăreşte doar o simplă replicare a lor. Dimpotrivă, prin modificarea caracteristicilor eşantionului şi, mai ales, prin modificarea bateriei (setului de itemi) vom ajunge la clarificări conceptuale esențiale. De exemplu, dacă un studiu nu reuşeşte să evidențieze foarte clar un anumit factor, care are importanță teoretică, putem reține itemii care au fost puternic saturați în factorul respectiv, putem elimina itemii care nu au dat satisfacție în această privință şi putem introduce alții, care par a măsura factorul respectiv.

Orice studiu care implică analiza factorială exploratorie rezidă în parcurgerea următorilor paşi:

(a) alegerea tipului de matrice care va fi analizată şi a metodei de extragere a factorilor;
(b) stabilirea numărului de factori ce trebuie extraşi;
(c) rotirea factorilor pentru a facilita interpretarea acestora (dacă este cazul);
(d) identificarea naturii factorilor şi interpretarea rezultatelor găsite.

Vom aborda detaliat fiecare aspect pornind de la următorul exemplu din psihologia educațională.

Didactogenia este acțiunea involuntară a cadrelor didactice soldată cu o serie de consecințe negative asupra elevilor. Între cele mai răspândite astfel de acțiuni, se regăsesc ridiculizarea elevului în public; favoritismul şi incorectitudinea în notare; respectiv agresivitatea verbală a dascălului. Consecințele acestor acțiuni asupra elevului pot fi grupate în: efecte pedagogice (ex. scăderea motivației şcolare, absenteism, eşec şcolar etc.); efecte psihologice (ex. stimă de sine scăzută, mutism electiv, manifestări anxioase, frustrare crescută); şi efecte somatice sau medicale (ex. dureri de cap, aciditate gastrică, tulburări intestinale etc.) (Poenaru, 1980; Poenaru şi Sava, 1998).

Vom analiza, în continuare, o scală menită să evalueze aceste efecte la elevii de liceu. Proba conține 15 itemi care necesită răspunsuri pe o scală de tip Likert cu răspunsuri notate de la unu la cinci. Ea a fost prezentată într-un studiu anterior, în care se urmăreau efectele atitudinii conflictuale, ostile a profesorilor asupra elevilor (Sava, 2002). Cercetarea a avut un eşantion de 946 de elevi, care trebuiau să evalueze un număr de 111 profesori. Astfel, fiecare profesor era caracterizat în medie de 8,52 elevi. Deoarece analiza factorială porneşte de la asumpția independenței dintre cazuri, am realizat o medie a răspunsurilor date de elevi pentru fiecare profesor caracterizat. Astfel, eşantionul total supus analizei factoriale a fost de 111 subiecți.

Itemii scalei sunt prezentați în tabelul 5.1., iar baza de date este analiza factorială exploratorie.sav. Pentru a realiza o analiză factorială de acest tip în SPSS se utilizează calea: „Analyze" – „Data reduction" – „Factor".

Instrucțiuni:

Indicați în ce măsură vi se potrivesc următoarele afirmații utilizând scala de notare:

1 – nu mi se potriveşte deloc; *4 – mi se potriveşte deseori;*

2 – mi se potriveşte rareori; *5 – mi se potriveşte în totalitate.*

3 – nu ştiu / indecis;

Tabelul 5.1. Itemii scalei de evaluare a efectelor psihologice, somatice şi pedagogice

ITEMI	
1. Îmi plac orele de X.	1 2 3 4 5
2. Dacă ar fi după mine aş chiuli de la ora de X.	1 2 3 4 5
3. Dacă toate disciplinele de la Bacalaureat ar conta la fel, aş alege să dau examen la X.	1 2 3 4 5
4. Mă plictisesc repede la orele de X.	1 2 3 4 5
5. Îmi place să învăț la X.	1 2 3 4 5
6. M-am simțit umilit de către profesor.	1 2 3 4 5
7. Profesorul îmi insuflă încredere în forțele proprii.	1 2 3 4 5
8. Mi-am imaginat moduri de răzbunare pentru suferința prilejuită de conduita profesorului.	1 2 3 4 5
9. Când mă gândesc la ora de X simt un gol în stomac.	1 2 3 4 5
10. Am uneori impresia că nu sunt în stare de nimic.	1 2 3 4 5
11. În general mă simt plin de energie la şcoală.	1 2 3 4 5
12. În timpul orei de X palmele îmi transpiră mai tare decât la alte ore.	1 2 3 4 5
13. După o zi de şcoală spatele (gâtul) meu devine încordat.	1 2 3 4 5
14. Mi s-a întâmplat să mi se facă rău la şcoală.	1 2 3 4 5
15. După ora de X mă simt obosit.	1 2 3 4 5

Unde X vizează disciplina predată de cadrul didactic evaluat.

5.1.3.1. Tipuri de matrice şi metode de extragere a factorilor

Această etapă este una pregătitoare. În contextul EFA există două tipuri de matrici: cea de corelație şi cea de covarianță. Aşa cum le spune numele, prima este o formă standardizată a celei de a doua. Matricea de corelații este modalitatea cel mai frecvent folosită datorită uşurinței interpretării relațiilor dintre variabile şi a saturației acestora în factorii extraşi. Metoda are şi un mic dezavantaj, în comparație cu matricea de covarianță, introducând o cantitate suplimentară de eroare, ca rezultat al standardizării. În cazul unor date măsurate pe scale foarte diferite ca amplitudine (ex. viteza de reacție exprimată în ms şi randamentul şcolar exprimat în medie şcolară) este binevenită standardizarea variabilelor înainte de a opta pentru matricea de corelații sau cea de covarianță. În alte cazuri, în care toții itemii folosesc aceeaşi scală de măsurare, nu este necesară standardizarea variabilelor. Pentru scala de față, vom păstra datele brute în vederea analizei şi vom opta pentru matricea de corelații, din rațiunile menționate mai sus.

Prin analiza factorială se încearcă găsirea unui „pattern" în matricea de corelații. Printr-un algoritm matematic sunt calculați anumiți parametri (ex. saturația factorilor, gradul de comunalitate etc.) care conduc la un model teoretic al matricei cât mai apropiat de cea observată în realitate. Dacă cele două matrici au „patternuri" de corelații similare, implicit valori reziduale scăzute, putem afirma că soluția găsită este adecvată. În cazul unei matrici reduse ca mărime, găsirea unui „pattern" s-ar putea face inspectând vizual corelațiile obținute. Într-o matrice complexă, acest lucru devine imposibil, analizei factoriale revenindu-i meritul de a sesiza eventualul „pattern" între corelații. În exemplul nostru, matricea de corelații obținută a fost:

Tabelul 5.2. Matricea de corelații dintre cei 15 itemi ai scalei

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. ed01	--														
2. ed02	**.60**	--													
3. ed03	**.73**	**.54**	--												
4. ed04	**.71**	**.64**	**.56**	**--**											
5. ed05	**.62**	**.53**	**.66**	**.50**	--										
6. ed06	**.55**	**.40**	**.35**	**.35**	**.36**	--									
7. ed07	**.53**	**.37**	**.40**	**.44**	**.51**	**.36**	--								
8. ed08	**.53**	**.50**	**.27**	**.38**	**.39**	**.73**	**.38**	--							
9. ed09	**.54**	**.41**	**.30**	**.39**	**.29**	**.66**	**.46**	**.61**	--						
10. ed10	**.27**	.19	.09	.18	.12	**.50**	.21	**.39**	**.48**	--					
11. ed11	.07	.15	-.06	.00	.10	.05	.09	.03	.01	**.22**	--				
12. ed12	**.32**	.20	.18	.16	.21	**.51**	**.22**	**.48**	**.64**	**.50**	-.06	--			
13. ed13	-.01	.04	-.04	-.07	.05	.05	-.08	.05	.02	.08	**.28**	-.02	--		
14. ed14	-.08	.16	-.07	.00	-.03	.02	.03	.06	.03	**.24**	**.27**	.05	**.39**	--	
15. ed15	**.53**	**.44**	**.34**	**.40**	**.37**	**.59**	**.36**	**.61**	**.67**	**.47**	.13	**.57**	.15	.15	--

Coeficienții îngroşați sunt semnificativi la $p < .01$

În privința metodelor de extragere a factorilor există o diversitate de alternative. Probabil, cea mai întâlnită metodă este analiza factorială a componentelor principale, care este opțiunea standard în SPSS. Atenție mare însă, între această analiză şi analiza componentelor principale, în sensul ei clasic, nu există un semn de egalitate. SPSS-ul nu oferă posibilitatea realizării metodei componentelor principale clasice. Ceea ce face SPSS-ul sub denumirea "principal component method" nu este altceva decât analiza factorială a componentelor principale. Diferența dintre cele două este totuşi minimă şi constă în faptul că opțiunea din SPSS conduce la o standardizare suplimentară după rotire, introducând o sursă suplimentară de eroare. Varianta clasică a analizei componentelor principale poate fi regăsită în PRELIS din cadrul programului statistic LISREL. O altă modalitate de extragere, întâlnită destul de frecvent, este

analiza factorilor principali (în engleză „principal axis factoring"). Ambele metode au avantajul că nu sunt foarte pretențioase în ce privește prezența unei distribuții multivariate normale, dar și dezavantajul că pot conduce la soluții relativ diferite, dacă sunt analizate pe baza matricei de corelații sau a celei de covarianță. Deși sensibilă la forma distribuției datelor, analiza factorială bazată pe estimarea verosimilității maxime (în engleză „maximum likelihood factor analysis") nu este afectată de tipul de matrice analizată. În plus, are marele avantaj de a testa gradul de semnificație statistică a saturației factorilor și a corelației dintre factorii extrași. Aceasta constituie o noutate în domeniul analizei factoriale exploratorii și poate fi valorificată utilizând programul CEFA (Comprehensive Exploratory Factor Analysis). Programul, realizat de cercetători de primă seamă în domeniu, poate fi descărcat gratuit de pe internet (Browne, Cudeck, Tateneni și Mels, 1998). SPSS-ul, deși conține posibilitatea extragerii factorilor cu această metodă, nu oferă avantajele prezente în CEFA. Singurul indicator prezent în SPSS, legat de gradul de potrivire a modelului, nu este de mare ajutor, fiind bazat pe o distribuție χ^2, deci, foarte sensibil la volumul eșantionului. Totuși, dacă veți folosi această metodă, veți extrage cel mai mic număr de factori de la care χ^2 devine nesemnificativ statistic, fapt ce înseamnă că nu mai există diferențe semnificative între matricea de corelații (covariabile) observată și cea estimată pe baza factorilor extrași.

5.1.3.2. Metode de selectare a numărului adecvat de factori

În stabilirea numărului de factori trebuie să realizăm o balanță între principiul parcimoniei (variabilele să fie influențate de un număr cât mai mic de factori) și principiul acurateții (obținerea unui model suficient de adecvat realității existente).

În ceea ce ne privește, am putea distinge două categorii de metode: statice (a priori) și dinamice. În prima categorie voi include metode ce utilizează criterii de selecție prestabilite.

Criteriul procentajelor de dispersie explicată

Un astfel de caz este criteriul procentajului de dispersie explicată (întâlnit în special în metoda componentelor principale). Un cercetător poate decide în avans o anumită cantitate de dispersie explicată pe care să o considere satisfăcătoare. Să presupunem, că el se oprește asupra unui procent de 70%. În această situație, el va selecta atâți eigenvectori (factori sau componente principale) câți sunt necesari pentru a explica acest procent. Dezavantajul major al metodei este dat de caracterul subiectiv al acestor repere a priori. Nimic nu ne oprește să decidem la fel de bine în favoarea unui procent de 60% din dispersie. Mai mult, în cazul analizei factoriale, multe dintre cercetările publicate se mulțumesc, în cazul științelor sociale, cu un procentaj de 40%.

În cazul problemei noastre, extragerea a trei factori comuni, care explică 63% din totalul dispersiei poate fi considerată mulțumitoare. Rezultatele oferite de SPSS 11.0 sunt prezentate în figura 5.2.

Merită să fie menționate câteva aspecte. Mai întâi, procentajul se stabilește pe baza valorilor eigenvectorilor extrași, denumite eigenvalue. În cazul utilizării matricelor de corelații, eigenvalue pentru fiecare variabilă inclusă în studiu este egală cu unu. Pe baza acestei informații putem observa puterea explicativă a fiecărui factor extras. Astfel, primul factor are un eigenvalue de 5,90 ceea ce înseamnă că puterea sa explicativă este aproape egală cu suma a șase dintre cei cincisprezece itemi ai scalei. Valoarea explicată prin suma celor trei eigenvalue corespunzătoare factorilor extrași se păstrează și după rotire. În situații în care efectuăm rotații ortogonale se păstrează și procentajul de 63%. Totuși, după o rotire oblică, precum în cazul problemei de față, procentajul explicativ al factorilor extrași nu mai este precizat, deoarece suma acestora depășește totalul dispersiei.

În sfârșit, putem desprinde valoarea explicativă a fiecărui factor, înainte și după rotire. Astfel, înainte de rotire, primul factor explica 39,34% din totalul dispersiei; cel de-al doilea, 13,53%, iar ultimul, 10,75%. Totalul acestora era de 63,62%. După rotația oblică, se observă o relativă egalitate între valoarea explicativă a primilor doi factori (eigenvalue de 4,81, respectiv 4,60), în timp ce al treilea factor are o putere explicativă mai redusă (1,80).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	5,901	39,341	39,341	5,901	39,341	39,341	4,816
2	2,030	13,530	52,871	2,030	13,530	52,871	4,605
3	1,614	10,757	63,628	1,614	10,757	63,628	1,802
4	,834	5,557	69,184				
5	,739	4,925	74,109				
6	,673	4,485	78,594				
7	,627	4,179	82,772				
8	,510	3,398	86,170				
9	,492	3,277	89,447				
10	,360	2,401	91,848				
11	,336	2,238	94,086				
12	,298	1,989	96,075				
13	,238	1,588	97,663				
14	,206	1,372	99,035				
15	,145	,965	100,000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Figura 5.2. *Output-ul SPSS 11.0 vizând numărul de factori extrași*
SPSS este o marcă înregistrată.

Criteriul lui Kaiser

Un alt criteriu, extrem de popular în selecția factorilor este regula lui Kaiser. Conform acesteia, vor fi selectați un număr de factori comuni egali cu numărul de eigenvalue care sunt mai mari decât unu. Adică, vor fi reținuți factorii care au o putere explicativă mai mare decât o singură variabilă, ținând cont că aceasta este compusă, nu numai din factori comuni, ci și din factori specifici și erori

de măsurare. Regula are o validitate acceptabilă. Totuşi, ea prezintă o serie de dezavantaje: poate fi aplicată doar în cazul analizei factoriale a matricelor de corelaţii, nu oferă criterii de decizie clare în cazul în care obţinem valori foarte apropiate de pragul critic unu (ex. eigenvalue = .99 sau 1.02), tinde uşor să supraestimeze numărul factorilor necesari. Totuşi, în cazul unor eşantioane mari, în situaţii cu mai puţin de 30 de variabile care toate au o comunalitate mai mare de .60, se dovedeşte a fi un indicator valid, la fel de eficient ca celelalte metode prezentate în continuare.

Pentru exemplul nostru se observă că primii trei factori îndeplinesc condiţia de a avea un eigenvalue mai mare decât unu (5,90; 2,03; şi 1,61), fapt ce indică o soluţie trifactorială.

Criteriul grafic al lui Cattel

Din categoria metodelor dinamice, ne vom opri asupra metodei reprezentării grafice a lui Cattell, cunoscută sub numele de scree plot şi asupra metodei analizei paralele propusă de Horn. Scree plot-ul este o reprezentare grafică a intensităţii eigenvalue pentru fiecare factor posibil a fi extras. În situaţia concretă este vorba de 15 factori posibili, dat fiind faptul că scala conţine 15 itemi. Aceştia sunt prezentaţi pe abscisă, în timp ce pe ordonată sunt trecute valorile eigenvalue. Principiul selecţiei ţine seama de acel număr de factori de la care soluţia nu se îmbunătăţeşte semnificativ. Grafic, acest lucru apare reprezentat sub forma unei cotituri ce indică ideea unei rupturi între situaţiile anterioare, în care se constată îmbunătăţiri semnificative ale soluţiei şi situaţiile ulterioare, în care soluţiile se ameliorează nesemnificativ. Graficul SPSS al problemei date este vizibil în figura 5.3.

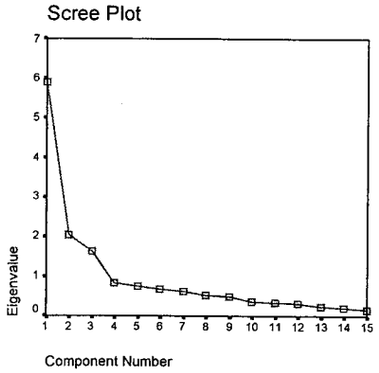


Figura 5.3. *Analiza numărului de factori ce trebuie extraşi pe bază de scree plot*
SPSS este o marcă înregistrată.

Se observă, pornind de la dreapta spre stânga, că linia începută în dreptul celui de-al cincisprezecelea factor continuă până la nivelul factorului patru. Primii

trei factori se află deasupra acestei linii, după acel punct de cotitură. În consecinţă, vor fi extraşi, ducând la o soluţie adecvată cu trei factori.

Metoda a condus, mai mereu, la rezultate valide. Dezavantajul utilizării ei apare în situaţii în care punctul de cotitură nu este atât de evident, fiind dificilă localizarea şi identificarea acestuia.

Metoda analizei paralele

Metoda analizei paralele se bazează, aşa cum arată numele, pe compararea intensităţii eigenvalue dintre datele observate şi un set de date aleatoare. Pe baza simulărilor se pot stabili valori aproximative eigenvalue în funcţie de numărul de subiecţi, numărul de variabile şi numărul de factori extraşi. În aceste condiţii, principiul de selecţie stipulează extragerea acelor factori care au eigenvalue mai mare decât cel corespondent, rezultat în urma analizei datelor aleatoare. O analiză a lui Zwick şi Velicer (1986) arată că această metodă dă rezultate valide, indiferent de condiţiile simulării. Totuşi, metoda are limite, cea mai importantă fiind dificultatea de a decide alegerea unui factor, în condiţiile în care există diferenţe infime între intensitatea eigenvalue obţinută din date reale şi cea corespondentă, rezultată în urma simulării. Din păcate SPSS-ul nu oferă direct posibilitatea realizării unei analize paralele.

În schimb, vom putea adapta ecuaţia de regresie dezvoltată de Allen şi Hubbard (1986, apud Johnson şi Wickern, 1998) la datele oferite de SPSS pentru a obţine această analiză într-un mod indirect. Ecuaţia de regresie a fost dezvoltată pentru a nu mai fi nevoiţi să realizăm simulări pe date randomizate şi are ca rezultat estimarea eigenvalue din date randomizate standardizate (ex. matrice de corelaţii). Regula de selecţie stipulează alegerea acelor factori ai căror eigenvalue este mai mare decât valoarea rezultată prin analiza datelor randomizate. Deşi metoda a fost dezvoltată pentru analiza componentelor principale, rezultatele sunt concludente şi în EFA.

Ecuaţia de regresie are forma:

$$\ln \lambda_k = a_k + b_k \ln(n-1) + c_k \ln\left[\frac{(p-k-1)(p-k+2)}{2}\right] + d_k \ln \lambda_{k-1} \quad (5.1)$$

unde: a, b, c şi d sunt nişte coeficienţi ce pot fi luaţi din tabel;
n este volumul eşantionului;
p este numărul de factori posibil a fi extraşi (numărul de variabile manifeste),
k este un indicator ce face referire la un anumit factor.
Se mai postulează ca lambda zero să fie egal cu 1.

Iată câteva valori ale coeficienților a, b, c şi d pentru soluții de până la opt factori.[1]

Tabelul 5.3. Valorile coeficienților ecuației de regresie

K	*a*	*b*	*c*	*d*	K	*a*	*b*	*c*	*d*
1	.9794	-.2059	.1226	.0000	5	-.2670	.0360	-.0024	1.0899
2	-.3781	.0461	.0040	1.0578	6	-.2632	.0368	-.0040	1.1039
3	-.3306	.0424	.0003	1.0805	7	-.2580	.0360	-.0039	1.1173
4	-.2795	.0364	-.0003	1.0714	8	-.2544	.0373	-.0064	1.1421

Dificultatea metodei este determinată de caracterul "inestetic" al ecuației, aspect contrabalansat de acuratețea rezultatelor obținute.

Vom calcula în scop demonstrativ, valorile eigenvalue obținute prin intermediul ecuației pentru primii doi factori. Reamintim că volumul eşantionului a fost de 111 profesori (n = 111), scala conținea 15 itemi (p = 15) şi că, prin convenție, λ zero este egal cu unu.

Pentru primul factor vom avea:

$$\ln \lambda_1 = .9794 - .2059 * \ln(111-1) + .1226 * \ln \frac{(15-1-1)(15-1+2)}{2} + .0000 * \ln 1 = .5811$$

Astfel, valoarea eigenvalue extrasă în cazul primului factor (vezi figura 5.2.) a fost de 5,90, fiind mai mare decât cea obținută pe baza datelor randomizate (1,78). În consecință, primul factor a trecut pragul de selecție şi va fi reținut.

Similar, putem calcula valoarea eigenvalue obținută din date randomizate şi pentru cel de-al doilea factor:

$$\ln \lambda_1 = -.3781 + .0461 * \ln(111-1) + .0004 * \ln \frac{(15-2-1)(15-2+2)}{2} + 1.0578 * .5811 = .471$$

$$\lambda_1 = e^{.471} = 1.599$$

Eigenvalue obținut din datele reale (2,03) este de asemenea mai mare decât cel obținut prin această ecuație (1.599). Prin urmare şi al doilea factor va fi reținut. Calculând intensitatea eigenvalue pentru al treilea şi al patrulea factor vom obține 1,46, respectiv 1,34. Compararea acestor valori cu cele obținute din datele reale indică reținerea celui de-al treilea factor (1,614), dar respingerea celui de-al patrulea (0,834). Aşadar, prin utilizarea unei adaptări a analizei paralele, soluția conține tot trei factori. Utilizatorii analizei paralele obişnuiesc să-şi prezinte rezultatele sub forma unui grafic asemănător celui din figura 5.4.

[1] Cei interesați de situații factoriale mai complexe pot consulta Allen şi Hubbard (1986, apud Johnson şi Wickern, 1998).

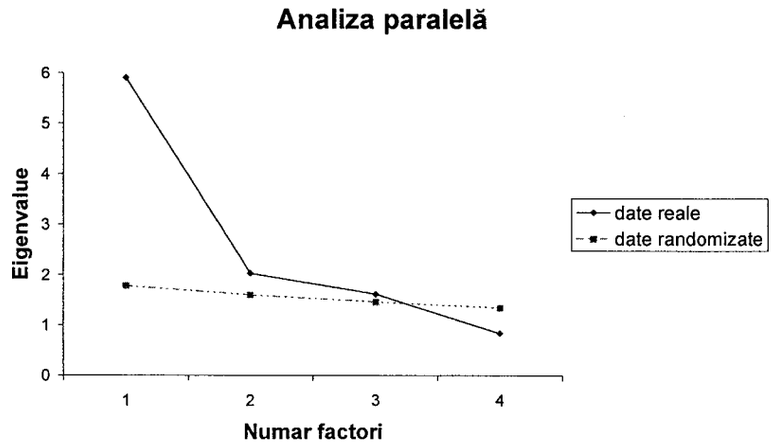


Figura 5.4. *Reprezentarea grafică a factorilor prin metoda analizei paralele*

În urma aplicării mai multor metode de selecție a factorilor, soluția favorită indică extragerea a trei factori. Într-un mod similar se procedează şi în cazurile de analiză a componentelor principale. Prin utilizarea a două sau a mai multor modalități de selecție ce oferă aceeaşi soluție, cresc semnificativ şansele ca aceasta să fie şi soluția corectă pentru eşantionul respectiv.

5.1.3.3. Analiza statistică a factorilor extraşi

Comunalitatea

După extragerea factorilor, vom observa gradul de comunalitate a fiecărei variabile şi saturația acestora pentru fiecare factor extras. Comunalitatea se referă la procentul din dispersia datelor care este explicat de acțiunea comună a factorilor reținuți şi se calculează pentru fiecare variabilă în parte. Cu cât comunalitatea este mai mare, cu atât acea variabilă este influențată mai mult de factorii extraşi. Similar, cu cât comunalitatea este mai mică, cu atât variabila respectivă este mai puțin influențată de acțiunea comună a factorilor extraşi. În plus, conform teoriei analizei factoriale prezentate anterior, dispersia care nu este explicată de comunalitate se presupune a fi legată de unicitate, adică de erori şi de factori specifici. Bunăoară, valorile comunalității mai mici de .50 arată că influența celor trei factori extraşi este mai mică decât dispersia datorată unicității. Prin urmare, itemii respectivi sunt mai puțin saturați în cei trei factori extraşi. În exemplul dat, se observă că itemii referitori la nivelul de energie fizică şi la gradul de încredere în forțele proprii sunt cel mai puțin explicați de cei trei factori extraşi. În schimb, itemul legat de preferința pentru obiectul de studiu şi cel privitor la prezența unor probleme gastrice au cel mai mare nivel de comunalitate. Altfel spus, ei sunt în cea mai mare măsură influențați de cei trei factori extraşi.

Tabelul 5.4. Saturația itemilor în cei trei factori extraşi

Etichetă itemi	SATURAȚIE ÎN FACTORI			Comunalitate
	1	2	3	
preferință obiect (1)	,837	-,311	,079	,804
chiul* (2)	,707	-,210	,314	,644
alegere examen (3)	,647	-,512	,154	,705
plictiseală* (4)	,686	-,389	,164	,649
învățare obiect (5)	,664	-,389	,253	,656
umilință suferită (6)*	,766	,237	-,220	,692
încredere percepută (7)	,621	-,203	,061	,431
răzbunare (8)*	,760	,198	-,156	,642
probleme gastrice (9)*	,775	,264	-,286	,752
lipsă încredere (10)*	,514	,557	-,098	,584
energie fizică (11)	,109	,332	,610	,493
transpirație (12)*	,582	,401	-,425	,681
dureri musculare (13)*	,047	,433	,590	,538
stare de rău (14)*	,082	,504	,569	,584
oboseală (15)*	,769	,309	-,053	,690

Notă: itemii notați cu asterix au fost recodați invers; în paranteze este trecut numărul de ordine al itemilor în scală

Saturația

Un alt nivel de analiză este legat de gradul de saturație a variabilelor în fiecare factor reținut. Astfel, se observă că itemul 1 este puternic saturat în primul factor şi mai puțin încărcat cu factorii doi şi trei. În schimb, itemul 11 pare a fi influențat mai ales de factorul trei şi prea puțin cu ceilalți factori. Din păcate, această soluție inițială este greu de interpretat, de cele mai multe ori datorită dualității saturațiilor. De exemplu, itemul 3 este influențat, în proporții relativ egale, de primii doi factori, iar itemul 12 pare a fi influențat într-o măsură relativ egală de toți cei trei factorii. Această stare de fapt nu este dorită, deoarece scopul primordial al analizei factoriale este identificarea factorilor extraşi. Acest lucru se poate face doar dacă structura saturației itemilor este *simplă*, ceea ce înseamnă ca fiecare variabilă să fie influențată predominant de un singur factor. Numai astfel vom putea identifica cu claritate natura factorilor extraşi.

Structura simplă

Pentru a obține o structură simplă se realizează o operație suplimentară de rotire a factorilor. Înainte de a explica pe îndelete această operație, merită să facem câteva precizări suplimentare. Comunalitatea, pentru fiecare variabilă, nu este altceva decât suma pătratelor fiecărei saturații factoriale. Astfel, dacă luăm itemul 1, observăm că suma pătratelor celor trei saturații este egală cu comunalitatea pentru itemul respectiv.

$$0.804 = (0.837)^2 + (-0.311)^2 + (0.079)^2$$

Valoarea comunalității se păstrează şi după rotire, chiar dacă saturațiile variabilelor în factorii respectivi se vor schimba. Într-un mod similar putem explica şi intensitatea eigenvalue, deşi, spre deosebire de comunalitate, valorile sale se modifică în urma rotirii. Dacă observăm pe verticală, gradul de saturație a fiecărei variabile într-un anumit factor vom putea obține nivelul eigenvalue. De exemplu, intensitatea primului factor poate fi obținută după formula:

$$5{,}90 = (0.837)^2 + (0.707)^2 + (0.647)^2 + \ldots + (0.769)^2$$

5.1.3.4. Rotirea factorilor

În urma selectării numărului adecvat de factori vom putea stabili saturația într-un anumit factor pentru fiecare item în parte. Acest fapt este necesar în interpretarea factorilor. De fapt, tabelul prezentat anterior se constituie, între anumite limite, într-o matrice a corelațiilor care există între fiecare item al scalei şi factorii extraşi. Ca în orice proces de corelație, cu cât saturațiile erau mai mari, cu atât exista o legătură mai puternică între factorul respectiv şi conținutul itemului în cauză. Au fost propuse mai multe praguri valorice ale gradului de saturație de la care să se considere că o anumită variabilă este încărcată în factorul respectiv. Cele mai întâlnite sunt pragurile de .30, .40 şi .50. Ce înseamnă aceste praguri dacă apelăm la interpretarea coeficienților de determinare corespondenți? Luarea în considerare a unei saturații de .30 indică o dispersie comună de numai 9% ($.30^2$) între factorul în cauză şi variabila respectivă. Acceptarea unui prag de .40 indică selectarea acelor itemi, ale căror dispersii sunt în proporție de 16% influențate de factorul în cauză. În sfârşit, acceptarea unui prag de .50 ca limită pentru a considera că un item este saturat în factorul respectiv, indică o dispersie comună de 25%. Desigur, aceste valori de corelație sunt dependente de volumul eşantionului, deoarece o dată cu creşterea acestuia vom obține valori tot mai mici ale coeficientului de corelație. În aceste condiții, autori precum Hair et al. (1998) propun câteva praguri de acceptat, care să ofere o putere statistică de .80, în condițiile unui $p < .05$. Astfel, în cazul a 100 de subiecți, pragul minim de la care putem spune că un factor influențează o anumită variabilă este .55, în timp ce un prag de .30 este suficient pentru un volum al eşantionului de cel puțin 350 subiecți. De reținut că aceste valori nu trebuie interpretate în termeni absoluți. Astfel, nu putem spune că un item nu este influențat de un anumit factor, dacă saturația sa este de .53, la un eşantion cu 100 de persoane. În schimb, putem aprecia că o saturație de .33 nu merită a fi luată în calcul, deşi aceeaşi valoare ar fi fost mai relevantă pentru un eşantion de 400 de persoane.

Analizați dinamica saturațiilor din tabelul anterior pentru cei 111 profesori evaluați. Se observă că cei mai mulți itemi sunt încărcați în primul factor (ex. toți itemii, cu excepția 11, 13 şi 14). Acest rezultat este normal pentru o soluție nerotită.

Merită, de asemenea, menționat că există o serie de itemi care nu au o structură simplă, par a fi influențați de acțiunea concomitentă a mai multor factori (ex. itemii 3, 10, 12, 13 şi 14). Acest rezultat îngreunează interpretarea factorială a soluției. De aceea, se obişnuieşte frecvent să se aplice rotirea factorilor pentru a elimina aceste probleme şi a obține structura cea mai simplă.

Criteriul structurii simple a fost definit de Thurstone. Ce înseamnă acest lucru? Din multitudinea de rotații posibile, cele care conduc la structuri simple vor fi mai uşor interpretabile şi vor avea o semnificație psihologică reală. Această structură simplă duce la obținerea unor saturații ridicate într-un singur factor şi redusă în ceilalți. De asemenea, numărul de itemi saturați în primul factor va scădea, reechilibrând, într-o anumită măsură, modul în care fiecare factor extras contribuie la explicarea dispersiei. Vom selecta patru din itemii care par a fi încărcați în mai mult decât un singur factor.

Tabelul 5.5. Itemi cu structură factorială complexă

alegere examen (3)	,647	-,512
plictiseală* (4)	,686	-,389
lipsă încredere (10)*	,514	,557
transpirație (12)*	,582	,401

Rotirea factorilor poate fi mai uşor înțeleasă prin apelul la reprezentarea grafică din figura 5.5. Grafic, factorii sunt reprezentați prin sistemul de axe, iar o variabilă situată în apropierea unei axe indică saturație crescută în acea dimensiune (factor). Înainte de rotire, poziția celor patru variabile era incertă pentru a desemna şi interpreta cei doi factori. Toate cele patru variabile erau situate aproximativ la jumătatea distanței dintre sistemul inițial de axe. Prin rotirea celor două axe, cu aproximativ 40° în direcția acelor de ceasornic, structura s-a simplificat (vezi liniile punctate). Astfel, se observă că itemii 3 şi 4 sunt foarte

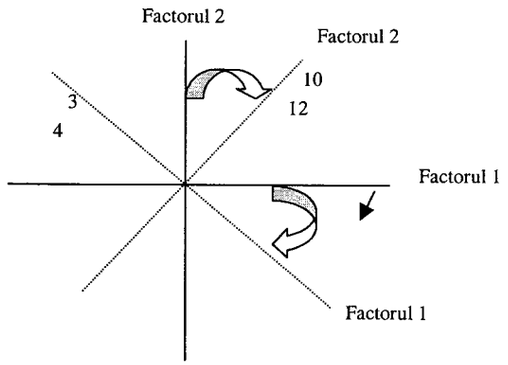


Figura 5.5. *Schematizarea procesului de rotire a factorilor*

saturați în factorul 1 şi prea puțin în cel de-al doilea. În schimb, pozițiile apropiate, ale itemilor 10 şi 12 de cea de-a doua axă indică o saturație puternică în acest factor şi o slabă saturare în primul.

În vederea obținerii structurii simple au fost dezvoltate mai multe modalități de rotire a factorilor. Cea mai importantă distincție între acestea împarte rotirile în două categorii: ortogonale şi oblice (Kachigan, 1991; Pedhazur şi Schmelkin, 1991). În cadrul rotirilor ortogonale, factorii sunt constrânşi să rămână necorelați, adică să formeze un unghi de 90°. Figura 5.5 prezintă o astfel de rotire, în care ambii factori sunt redirecționați spre dreapta cu aproximativ 40°. Acest fapt menține unghiul drept dintre cei doi factori.

În schimb, rotirile oblice presupun redirecționarea factorilor cu intensități diferite. Acest lucru face ca după rotire, unghiul dintre sistemul de axe să nu fie de 90°. Transpusă în termeni statistici, această informație precizează că factorii extraşi corelează între ei, nefiind independenți unul de celălalt. Figura 5.6 prezintă grafic diferența dintre cele două modalități de rotire.

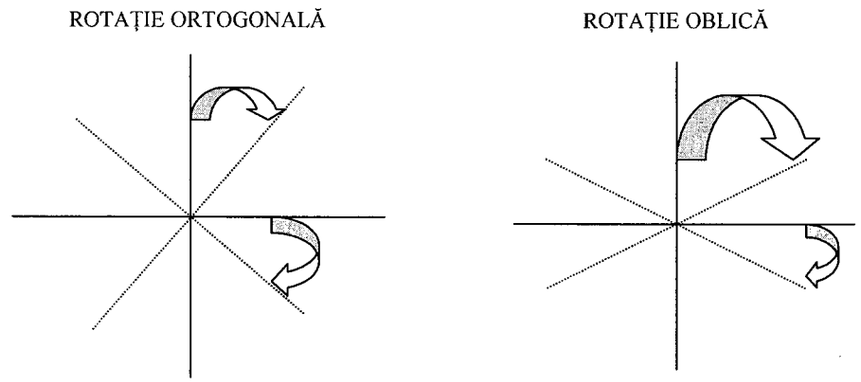


Figura 5.6. *Tipuri de rotiri ale factorilor*

Alegerea dintre cele două tipuri de rotire este dependentă de cunoştințele pe care le avem despre variabile şi despre presupuşii factori extraşi. Dacă credem că există corelații între dimensiunile latente extrase, ar trebui să optăm pentru rotiri de tip oblic. Dacă considerăm că factorii extraşi nu trebuie să coreleze, vom alege rotirea ortogonală. Evident că alegerea greşită a tipului de rotire factorială se soldează cu obținerea unei soluții îndepărtate de cea mai simplă structură, fapt ce îngreunează interpretarea factorilor.

În cazul nostru, obiectivul specificat era de a distinge între efectele educaționale, psihologice şi somatice rezultate în urma didactogeniei. Ne aşteptăm că aceste dimensiuni să fie independente, adică să nu coreleze deloc? Puțin probabil. Bunăoară, elevii afectați de comportamentele inadecvate ale profesorilor

pot suferi atât în plan somatic (ex. "După ora de X mă simt obosit".), cât şi în plan educaţional (ex. "Dacă aş avea posibilitatea, aş chiuli de la ora de X".). Aceste efecte se pot întrepătrunde, fiind probabilă existenţa unei corelaţii între cele două seturi de efecte. Lucrurile ar fi stat diferit, dacă ne-am fi propus să realizăm o probă de evaluare a flexibilităţii şi am fi dorit să păstrăm câţiva itemi într-o subscală de minciună. Informaţiile teoretice în domeniu nu specifică existenţa vreunei relaţii între flexibilitate şi scala de minciună. În aceste condiţii, s-ar fi impus utilizarea unei rotiri ortogonale pentru a obţine o structură cât mai simplă.

Dincolo de această distincţie esenţială, putem identifica mai multe modalităţi de rotire a factorilor (Garson, 2000; Field, 2000).

Varimax

Sub acest nume se găseşte cea mai utilizată rotire factorială de tip ortogonal. Ea îşi propune să minimizeze numărul de variabile care au o saturaţie crescută într-un anumit factor. Aceasta duce aparent la împărţirea itemilor: o parte având saturaţii crescute în acel factor, iar cealaltă parte, o saturaţie scăzută. Această rotire este recomandată în toate situaţiile în care decidem să rotim ortogonal factorii şi reprezintă opţiunea standard în SPSS.

Quartimax

Este o rotire ortogonală alternativă. Ea îşi propune să minimizeze numărul de factori necesari pentru a explica fiecare variabilă. Tehnica este mai puţin folosită deoarece tinde să determine majoritatea variabilelor să fie saturate într-un singur factor. Totuşi, în situaţiile în care teoria afirmă existenţa unui factor general unitar (precum factorul g din domeniul inteligenţei) este recomandată această rotire.

Direct oblimin

Aceasta este o rotire oblică flexibilă, deoarece permite factorilor să coreleze la intensităţi variate, prin modificarea valorii lor delta. În SPSS, delta este prestabilit la valoarea zero, ceea ce înseamnă că nu permite corelaţii ridicate între factori. O astfel de valoare duce frecvent la valori absolute ale coeficienţilor de corelaţie între factori cuprinse între 0 şi .50. Delta poate fi modificat între valorile –0.8 şi 0.8. Valorile pozitive permit corelaţii ridicate între factori, fiind posibile şi valori ale coeficienţilor de corelaţie de .80. Similar, valorile delta negative permit corelaţii foarte mici, coeficienţii de corelaţie fiind inferiori celor obţinuţi dacă păstrăm delta egal cu zero. Desigur că valoarea delta va fi stabilită tot din raţiuni teoretice. Totuşi, pentru cele mai multe situaţii, valoarea standard (delta = 0) poate fi păstrată. Această rotire este cunoscută şi sub numele de rotire directă quartimin, fiind cea mai întâlnită rotire din categoria celor oblice.

Promax

În fine, Promax este o altă procedură de rotire oblică, mai puţin utilizată datorită pretării sale mai ales la eşantioane foarte mari de subiecţi. Toate aceste tipuri de rotiri sunt prezente şi între opţiunile SPSS.

Rezultatele obţinute în urma rotirii factorilor sunt structurate în două matrici finale: matricea structurii factoriale (în SPSS: structure matrix) şi matricea modelului factorial (în SPSS: factor matrix). De remarcat că aceste matrici sunt

identice ca valori obţinute în cadrul rotaţiilor ortogonale, caz în care SPSS-ul oferă o singură matrice (în engleză „rotated component matrix" sau „rotated factor matrix").

Matricea structurii factoriale şi matricea modelului factorial

Matricea structurii factoriale conţine corelaţiile obişnuite (de ordinul zero) dintre variabilele originale şi factorii rotiţi. Cu cât o corelaţie este mai mare, cu atât variabila respectivă este mai strâns legată de factorul respectiv. În situaţii în care factorii sunt corelaţi între ei, valoarea corelaţiei dintre un item şi un anumit factor este mediată (influenţată) de relaţia dintre ceilalţi factori extraşi şi itemul respectiv. În schimb, matricea modelului factorial conţine ponderări (în engleză „weights") utilizate în determinarea scorului factorial. Acestea sunt corespondentul coeficienţilor standardizaţi beta din ecuaţia de regresie. Ponderările indică influenţa "pură" a unui factor asupra variabilei respective, eliminând influenţa celorlalţi factori. Cu alte cuvinte, ponderările indică valorile coeficienţilor de corelaţie semiparţială în situaţii de standardizare a datelor. Cele două matrici au valori similare în condiţiile în care factorii extraşi nu corelează între ei. Lipsa corelaţiei dintre factori este tipică cazurilor în care soluţiile factoriale obţinute nu au fost rotite sau au suferit rotiri de tip ortogonal. Din păcate, asemănările dintre ele creează confuzii pentru unii cercetători, matricea modelului factorial fiind interpretată greşit, ca o matrice de corelaţii dintre variabilele şi factorii extraşi (Pedhazur şi Schmelkin, 1991; Kline, 1993).

În cazul rotirilor oblice, SPSS-ul oferă, complementar, o matrice de corelaţii între factorii extraşi, pentru a sesiza structura şi intensitatea relaţiilor dintre aceştia. În cazul efectelor psihologice, educaţionale şi somatice din cazul prezentat anterior, SPSS-ul oferă următoarea matrice de corelaţii între factori:

Component Correlation Matrix

Component	1	2	3
1	1,000		
2	**,404**	1,000	
3	,010	,110	1,000

Valoarea ingrosata este semnificativa statistic la p < .001

Figura 5.7. *Output SPSS cu matricea corelaţiilor dintre factori*
SPSS este o marcă înregistrată.

În figura 5.8 sunt prezentate matricile de bază oferite de SPSS. În cele două se observă o asemănare în structură după rotirea oblică a celor trei factori extraşi. Matricea modelului factorial indică o grupare a itemilor 1, 2, 3, 4, 5 şi 7 după primul factor, a itemilor 6, 8, 9, 10, 12 şi 15 după al doilea factor, respectiv a

itemilor 11, 13 şi 14 după al treilea factor. Matricea structurii factoriale, ne arată corelațiile dintre fiecare item şi factorii în cauză.

Pattern Matrix[a]

	Component		
	1	2	3
ED03	**,868**	-,099	-,100
ED05	**,831**	-,059	,048
ED04	**,803**	,007	-,028
ED01	**,797**	,200	-,057
ED02	**,749**	,059	,194
ED07	**,572**	,167	-,026
ED12	-,145	**,877**	-,144
ED09	,147	**,803**	-,085
ED10	-,153	**,756**	,216
ED06	,192	**,738**	-,042
ED15	,217	**,693**	,139
ED08	,248	**,669**	-,007
ED14	-,069	,074	**,753**
ED13	-,028	-,009	**,734**
ED11	,098	-,061	**,702**

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 8 iterations.

Structure Matrix

	Component		
	1	2	3
ED01	**,878**	,516	-,027
ED03	**,827**	,240	-,103
ED05	**,807**	,282	,050
ED04	**,805**	,328	-,019
ED02	**,775**	,383	,208
ED07	**,639**	,395	-,002
ED09	,470	**,852**	,004
ED06	,490	**,811**	,041
ED12	,208	**,803**	-,049
ED15	,498	**,796**	,218
ED08	,518	**,768**	,070
ED10	,154	**,718**	,298
ED14	-,032	,129	**,760**
ED13	-,024	,060	**,732**
ED11	,081	,056	**,696**

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalizatior

Figura 5.8. *Matricea modelului factorial (în stânga) şi matricea structurii factoriale (în dreapta)*
SPSS este o marcă înregistrată.

5.1.3.5. Interpretarea naturii factorilor

Pornind de la matricea structurii factoriale rezultată în urma rotirii factorilor vom urmări să interpretăm factorii extraşi. Din ambele matrici ale figurii 5.8. se observă că primul factor este prezent în itemul 7 şi în primii cinci itemi. Corelația acestor variabile cu factorul respectiv variază între .639 şi .878. Cu cât corelația este mai mare cu atât variabila respectivă are un rol mai mare în interpretarea factorului. Se pune întrebarea generală, ce au în comun aceste variabile saturate puternic în primul factor? În plus, ce diferențiază acest set de variabile de celelalte care nu au obținut saturații ridicate în factorul respectiv? Din analiza de conținut (vezi matricea modelului factorial din figura 5.8. şi tabelele 5.1 sau 5.4) reiese că cei şase itemi privesc efectele în plan pedagogic, ei fiind strâns legați de motivația şcolară. Astfel, elevii afectați în această direcție nu sunt atraşi de orele respective, absentează de la acele ore atunci când au ocazia, nu învață la disciplina respectivă şi evită contactul cu aceasta. La celălalt capăt, se află elevii care nu sunt afectați în plan pedagogic, cei cărora le face plăcere să învețe şi sunt atraşi de disciplina respectivă, participând activ la orele respective.

Într-o manieră similară se procedează la analiza şi interpretarea celorlalți factori. Astfel, factorul doi pare a fi strâns legat de itemii 6, 8, 9, 10, 12 şi 15. Corelațiile acestor itemi cu factorul doi sunt foarte ridicate, variind între .718 şi .852. Analiza de conținut a acestor itemi evidențiază structura dublă, implicând efecte psihologice (ex. "M-am simțit umilit de către profesor".) şi efecte somatice (ex "După ora de X mă simt obosit".). În consecință, acest factor poate fi etichetat sub denumirea "efecte psihologice şi somatice" ale didactogeniei. În sfârşit, cel de-al treilea factor este strâns corelat cu itemii 11, 13 şi 14, valorile coeficienților de corelație pentru aceştia fiind cuprinse între .696 şi .760. Se remarcă, din conținutul itemilor, că toți trei fac referire la o stare somatică generală a elevilor la şcoală, fără a fi specifici unui anumit profesor (ex. "După o zi de şcoală gâtul meu devine încordat".).

Etichetarea factorilor

Aşadar, interpretarea factorilor este un proces subiectiv bazat pe matricea structurii factoriale sau pe matricea modelului factorial. Denumirile factorilor au la bază o bună cunoaştere teoretică a domeniului. Totuşi, pot exista cazuri în care etichetarea factorilor nu se face uşor. În astfel de situații, cercetătorii tind să creeze noi etichete pentru aceştia. Două exemple clasice în psihologie sunt date de denumirile celor 16 factori primari din teoria personalității a lui Cattell sau de personalitățile de tip A sau B din cercetările asupra stresului desfăşurate de Friedman.

După etichetarea itemilor se poate studia matricea corelațiilor dintre factorii extraşi. În problema de față, se observă că singura corelație semnificativă statistic este cea obținută între primul şi al doilea factor (r = .40), adică între efectele pedagogice şi efectele psihosomatice. Acest aspect pare normal, deoarece comportamentele inadecvate ale cadrelor didactice pot exercita influențe negative în planuri multiple. De asemenea, prin lipsa corelației dintre efectele pedagogice şi efectele somatice generale, respectiv dintre efectele psihosomatice şi efectele somatice generale, se constată o bună discriminare şi validare a faptului că primii doi factori măsoară efectele declanşate de conduita unui anumit cadru didactic, fiind specifice interacțiunii particulare dintre un anumit dascăl şi clasa de elevi.

5.1.3.6. Ameliorarea structurii factoriale

De multe ori, o dată cu interpretarea naturii factorilor, cercetătorii tind să finalizeze procesul analizei factoriale. Recomandarea mea este să nu ne oprim la acest nivel. Următorul pas constă în evaluarea posibilităților de modificare sau îmbunătățire a scalei. Se vor urmări câteva aspecte, cele mai importante fiind gradul de comunalitate şi de saturație al fiecărei variabile incluse în studiu.

Cu cât gradul de comunalitate este mai mare, cu atât variabila respectivă este influențată mai mult de factorii comuni extraşi. Spre exemplu, o comunalitate de .804 a itemului "Îmi plac orele de X." indică faptul că dispersia variabilei

respective este influențată în proporție de 80,4% de acțiunea factorilor comuni şi doar un procent de 19,6% se datorează unicității itemului, adică erorilor de măsurare şi factorilor specifici. O valoare a comunalității sub .50 indică faptul că itemul respectiv este influențat în mai mare măsură de aspecte unice, decât de acțiunea factorilor comuni extraşi. Totuşi, o variabilă care are o comunalitate mai redusă (ex. .40) poate fi păstrată în scală, dacă este încărcată într-un singur factor comun şi ajută la definirea acestuia. Această regulă nu se aplică în cazul comunalităților foarte scăzute (c < .30).

Rațiunile acestei decizii pot fi înțelese uşor prin următoarea demonstrație. Fie o scală cu doi factori, în care doi itemi au, ambii, o comunalitate de .40 şi grade diferite de saturație după o rotire ortogonală.

$$(1)\ .40 = (.62)^2 + (-.12)^2$$
$$(2)\ .40 = (.45)^2 + (.44)^2$$

În cazul primului item, se observă clar saturația în primul factor (.62) şi lipsa corelației dintre item şi factorul al doilea. Într-o astfel de situație, în ciuda unei comunalități relativ scăzute, putem păstra variabila respectivă deoarece oferă un bun ajutor în definirea primului factor. În schimb, în al doilea caz am putea exclude sau înlocui cel de-al doilea item deoarece, pe lângă comunalitatea scăzută, nu oferă nici claritate în interpretarea factorilor datorită saturațiilor egale în cei doi factori (ex. 45, respectiv .44). Acest criteriu al clarității interpretării factorilor poate sta la baza eliminării sau modificării itemilor care, deşi au un grad ridicat de comunalitate, nu ajută la interpretarea factorilor.

Aceste valori discutate ale comunalităților sunt valabile pentru analiza componentelor principale, sub forma ei clasică sau în cadrul analizei factoriale a componentelor principale oferită de SPSS. În cazul alegerii unei alte metode de extracție a factorilor, vom obține comunalități mai mici, de aceea pot fi acceptate şi valori mai reduse (ex. c = .20, în cazul analizei factoriale a axelor principale). Această situație se explică prin apelul la teorie. În cazul analizei componentelor principale, prin comunalitate se înțelege dispersia explicată ce include atât factori comuni, cât şi factori specifici. Sensul tradițional al comunalității se regăseşte în analiza factorială exploratorie. Aici, prin comunalitate se înțelege influența comună asupra dispersiei datelor exercitată de acțiunea simultană a factorilor comuni găsiți. Prin urmare, factorii specifici sunt excluşi din comunalitate, formând, împreună cu erorile de măsurare, unicitatea dispersiei.

Un alt motiv de înlocuire a unor itemi din scală este dat de apariția unor factori care în prealabil nu erau doriți. Acest rezultat este dependent în mare măsură şi de conținutul itemilor introduşi în analiză. În exemplul nostru, intenția a fost de a construi o scală care să evidențieze efectele comportamentului inadecvat al cadrelor didactice asupra elevilor. Itemii 11, 13 şi 14, care au constituit al treilea factor - "efecte somatice generale" -, nu pot fi incluşi în această categorie, deoarece nu corelează cu modul de comportare a dascălilor. Itemii țin mai degrabă

de starea generală de sănătate a elevului şi de particularitățile şcolarizării în general. În consecință, un studiu ulterior poate conduce la eliminarea acestor itemi, în condițiile în care dorim să ne concentrăm pe efectele declanşate de un anumit cadru didactic.

Reversul paragrafului precedent constă în neapariția unor factori care erau aşteptați în prealabil. În cazul de față, ne aşteptam să obținem o diferențiere mai clară între efectele psihologice şi cele somatice. În schimb, acestea au corelat puternic între ele, conducând la extragerea unui singur factor. În astfel de situații, se recomandă introducerea unor itemi suplimentari în cercetările ulterioare pentru a ameliora instrumentul de măsurare. Criteriul de selecție utilizat în introducerea acestor itemi ține de testarea structurii unilaterale sau bifactoriale a itemilor respectivi. Astfel, vom putea introduce doi itemi suplimentari, unul care face trimitere la efecte în plan somatic (ex. "În timpul orelor de X mă apucă durerile de cap."), iar altul care se referă la efecte în plan psihologic (ex. "Mă simt descurajat să-mi exprim punctele de vedere în timpul orei de X."). În măsura în care introducerea acestor itemi nu va diferenția între efectele psihologice şi cele somatice, am putea conchide că o structură comună este mai potrivită în acest caz. O regulă uzuală specifică precizează necesitatea existenței a cel puțin trei itemi pentru fiecare factor extras.

În concluzie, un studiu explorator ulterior poate urmări ameliorarea scalei de măsurare a efectelor nedorite prin introducerea următoarelor modificări:

(1) eliminarea itemilor 11, 13 şi 14 care nu fac referire la efectele declanşate de conduita unui anumit dascăl, ci măsoară o stare somatică generală;
(2) introducerea a doi itemi suplimentari pentru a vedea în ce măsură se poate diferenția între efectele psihologice şi cele somatice, adică pentru a determina cât de adecvată este extragerea unui singur factor pentru cele două tipuri de efecte.

5.1.3.7. Aspecte metodologice complementare

În analiza factorială merită să fie trecute în revistă câteva aspecte care influențează rezultatele demersului statistic. Principalele probleme ce trebuie discutate se referă la volumul şi caracteristicile eşantionului, la selectarea variabilelor manifeste dintr-un domeniu şi la condiții complementare în analiza factorială.

Volumul şi caracteristicile eşantionului

Analiza factorială este o tehnică pretențioasă în ceea ce priveşte numărul de subiecți necesari. Minimul acceptat pentru o astfel de analiză este de 100 de subiecți, însă este preferabilă utilizarea unor eşantioane de peste 200 de subiecți. În plus, volumul eşantionului este dependent de numărul de variabile supuse analizei. Se recomandă ca raportul număr de subiecți per număr de variabile să fie cel puțin egal cu cinci. Astfel, pentru o scală cu 30 de itemi, ar fi de dorit un minim de 150 de subiecți. Acest raport nu trebuie însă interpretat în mod rigid pentru cazurile în care volumul eşantionului devine destul de mare. Astfel, putem efectua

o analiză factorială adecvată pe un eşantion de 400 de subiecţi, chiar dacă numărul de variabile (itemi) este de 130. Mai rămâne să reamintim doar că volumul eşantionului determină, în anumită măsură, valorile de la care considerăm că un anumit item este saturat de factorii extraşi.

În ce priveşte caracteristicile eşantionului, soluţia ideală ar fi selectarea unui lot de persoane de o omogenitate medie. Un eşantion prea omogen poate induce o soluţie factorială greşită. Kline (1993) exemplifică foarte bine această stare de fapt prin situaţia următoare. Să presupunem că un cercetător doreşte să identifice care sunt factorii care afectează reuşita academică şi utilizează un lot de subiecţi de elită, studenţi ai Universităţilor Oxford şi Cambridge. Datorită omogenităţii lotului, am putea avea surpriza de a constata că inteligenţa nu joacă un rol important în reuşita academică, deoarece majoritatea studenţilor de la aceste universităţi au o inteligenţă superioară. Dacă eşantionul de subiecţi ar include studenţi de la toate universităţile britanice am avea mari şanse să demonstrăm că inteligenţa joacă un rol important în reuşita şcolară. Un exemplu revelator este faptul că foamea nu a fost găsită ca fiind un factor motivaţional fundamental într-o cercetare a lui Cattell şi Child (apud Kline, 1993), deoarece subiecţii erau proveniţi din pătura medie şi superioară a unor ţări foarte dezvoltate.

La polul opus se regăseşte selectarea unui eşantion foarte eterogen. În astfel de situaţii, apare riscul de a extrage o serie de factori irelevanţi pentru obiectivele urmărite, mulţi factori fiind rezultatul unor caracteristici demografice diferenţiatoare. Într-o asemenea situaţie se recomandă utilizarea analizei factoriale pentru fiecare „sub-eşantion" relevant.

Testarea gradului de adecvare a eşantionului se poate face prin metoda KMO (Kaiser-Meyer-Olkin) prezentă în SPSS Factor - Descriptives. Acesta poate lua valori între 0 şi 1. Dacă valorile obţinute sunt sub .50, eşantionul nu este adecvat pentru a realiza o analiză factorială. Valorile între .50 şi .70 sunt satisfăcătoare, cele peste .70 sunt adecvate, iar cele peste .80 sunt excelente pentru a aplica analiza factorială pe eşantionul respectiv (Field, 2000). Gradul de adecvare a eşantionului pentru fiecare variabilă în parte poate fi testat prin obţinerea unei anti-imagini, opţiune disponibilă în SPSS în aceeaşi secţiune. Matricea obţinută va avea, printre altele, o serie de valori pe diagonala principală, fiind necesare valori mai mari de .50 pentru fiecare variabilă în parte. Neîndeplinirea acestui criteriu poate duce la eliminarea itemului respectiv din analiză.

Selectarea variabilelor supuse analizei factoriale

Teoria analizei factoriale stipulează selectarea unui grup reprezentativ de variabile pentru domeniul de interes. Selectarea unor variabile care nu fac parte din domeniul respectiv se soldează cu obţinerea unor factori neaşteptaţi sau irelevanţi. Cei trei itemi din scala prezentată pe întreg parcursul acestui capitol pot fi incluşi în această categorie, deoarece ei nu se referă la domeniul efectelor produse de comportamentul unui anumit cadru didactic în particular. De asemenea, selectarea

unui număr de itemi subreprezentativi pentru domeniul de interes afectează conţinutul factorului respectiv. Pentru a clarifica această ultimă situaţie voi apela la teoria de personalitate Big-Five a lui Costa şi McCrae (1992). Unul dintre domeniile personalităţii este nivelul de agreabilitate. Acest domeniu conţine şase faţete reprezentative: încrederea (are încredere în alţi oameni), sinceritatea (spune ce are pe suflet), altruismul (este disponibil spre ajutorarea celorlalţi), conformismul (evită conflictele interpersonale), modestia (nu se comportă în mod arogant), şi implicarea afectivă (receptivitate la apeluri umanitare, opusul fiind persoanele care se consideră realiste şi iau decizii bazate pe o logică rece).

Pentru a măsura agreabilitatea este important să fie prezenţi itemi din toate aceste faţete. Dacă vom utiliza doar itemi care măsoară nivelul de altruism, vom obţine un factor care nu va putea fi etichetat "agreabilitate". În schimb, cel mai probabil, factorul va fi foarte distinct şi va avea o consistenţă internă ridicată. Validitatea de construct a agreabilităţii, măsurată exclusiv prin faţeta altruismului, va dovedi ulterior caracterul eronat al etichetării factorului prin acest domeniu general, în condiţiile în care celelalte faţete sunt absente.

Pornind de la acest aspect, se cuvine să precizăm că în analiza factorială se vorbeşte de două categorii de factori comuni: primari (rezultaţi din corelaţiile itemilor supuşi analizei factoriale) şi secundari (rezultaţi din corelaţiile factorilor primari selectaţi). Această organizare ierarhică se poate observa şi în cazul dimensiunilor personalităţii din Big Five şi nu numai. Astfel, agreabilitatea este din punct de vedere statistic un factor secundar bazat pe şase factori primari: încrederea, sinceritatea, altruismul, conformismul, modestia şi implicarea afectivă. Fiecare dintre aceştia au la bază itemi specifici care sunt măsuraţi direct prin afirmaţiile prezente în inventar.

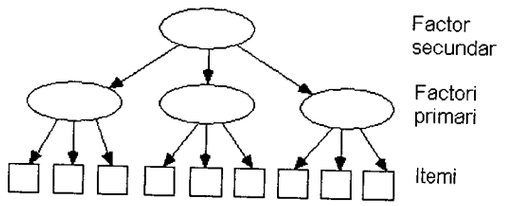


Figura 5.9. *Relaţia dintre itemi, factori primari şi factori secundari*

Asumpţii complementare ale analizei factoriale exploratorii

În această categorie vom prezenta câteva metode de testare a măsurii în care este potrivită utilizarea analizei factoriale.

O primă asumpţie priveşte *multicoliniaritatea* variabilelor manifeste. Pentru a efectua o analiză factorială semnificativă ar fi bine ca asocierile dintre variabile să nu fie foarte mari (ex. $r > .90$), deoarece există riscul de a măsura acelaşi lucru. SPSS-ul oferă posibilitatea de a calcula determinantul matricei de corelaţii în cadrul

secțiunii "Descriptives" din analiza factorială. Dacă valoarea acestuia este mai mică decât 0,00001, putem diagnostica o situație de multicoliniaritate. Într-o asemenea situație, va trebui să observăm care coeficient de corelație este mai mare de .80 sau .90 şi să eliminăm una dintre variabile, deoarece există şansa ca ele să măsoare acelaşi lucru.

O situație, oarecum opusă, este cazul în care variabilele nu corelează între ele, prin urmare nu sunt afectate de factori comuni. Testul de sfericitate al lui Barlett, prezent în "Descriptives", SPSS, testează o asemenea ipoteză: în ce măsură avem de a face cu o matrice de identitate (corelațiile dintre variabile, cu excepția diagonalei principale sunt zero)? Un rezultat semnificativ statistic elimină această posibilitate şi oferă şansa realizării analizei factoriale. Un rezultat nesemnificativ indică faptul că analiza factorială nu este potrivită pentru a fi aplicată. Din păcate, numai situațiile clare sunt diagnosticate, deoarece testul tinde, de regulă, a fi semnificativ datorită numărului mare de subiecți dintr-un eşantion.

Toate aceste teste preliminarii au fost efectuate şi pentru scala de față, rezultatele indicând o pretare foarte bună a variabilelor şi a eşantionului chestionat pentru a fi analizate factorial.

În sfârşit, analiza factorială poate fi aplicată în situațiile în care variabilele sunt măsurate prin scale de tip interval sau raport, iar în cel mai rău caz, prin scale de tip ordinal. Nu este recomandată analiza factorială clasică a datelor de tip categorial (ex. răspunuri de tip da sau nu), deoarece coeficienții de corelație phi nu pot lua valori pe întreaga gamă de la –1 la +1, iar relația dintre variabile nu poate fi liniară. În această situație se recomandă utilizarea altor metode de analiză factorială, precum "Categorical Principal Components" din pachetul adițional SPSS Categories 11.0 (Meulman şi Heiser, 2001) ori analiza claselor latente (Finch şi West, 1997).

Totuşi, printr-un artificiu tehnic se poate aplica analiza factorială exploratorie, în cazul datelor dihotomice, dacă se renunță la utilizarea itemului ca unitate de măsură, în favoarea scalei ori a unui pachet de itemi. Printr-un pachet de itemi se înțelege gruparea mai multor itemi şi analizarea lor împreună. Astfel, o scală care avea în prealabil 40 de itemi cu răspunsuri dihotomice de tip „Da" sau „Nu" poate fi analizată factorial în stil clasic după o grupare prealabilă a itemilor în 10 pachete a câte patru itemi fiecare. În acest fel, răspunsurile subiecților nu mai sunt notate binar („1 = Da", „0 = Nu"), ci pot lua valori de la 0 (în cazul în care subiecții răspund „Nu" la toți itemii dintr-un pachet) la 4 (dacă subiecții răspund „Da" la toate afirmațiile din pachet). Metoda,care a fost propusă de Cattell, iar într-o variantă modificată de Comrey (apud Kline, 1993), are două limite. Prima se referă la caracterul subiectiv de alegere a itemilor care vor fi grupați în acelaşi pachet, iar a doua limită vizează caracterul factorilor rezultați în urma unei astfel de analize. Aceşti factori sunt considerați de mulți autori ca fiind factori de ordinul doi şi nu factori primari.

5.1.4. Exemplu de analiză factorială exploratorie - cazul inteligenței

În cele ce urmează, voi prezenta mult mai succint paşii unei analize factoriale pornind de la o serie de probe ce măsoară inteligența. Baza de date include rezultatele de la nouă probe aplicate de Thurstone şi Thurstone în 1941 (apud Tucker şi MacCallum, 1997), pe un lot de 710 elevi din clasele a VII-a şi a VIII-a. Analiza factorială s-a realizat fără a avea acces la baza de date brute, pe baza matricei de corelații prezentate mai jos. SPSS-ul oferă această posibilitate prin sintaxă. Iată o scurtă descriere a celor nouă teste:

(1) probe de adunare – include itemi ce presupun adunarea pe coloane;
(2) probe de înmulțire – include itemi de înmulțire a două numere;
(3) proba seriei de numere – include itemi ce constau într-un şir de numere care se multiplicau cu trei, cum ar fi seria: 1, 3, 9, 27, ...; subiecții trebuiau să identifice numărul următor;
(4) proba figurilor – include o serie de litere sau semne care au fost rotite sau reflectate în oglindă; sarcina este de a identifica care figuri dintr-o serie sunt identice cu prima figură;
(5) proba cărților – include itemi sub forma unor cărți de joc reprezentând diverse figuri geometrice, rotite sau reflectate în diverse moduri; sarcina era de a găsi care cărți pot fi rotite pentru a se potrivi cu figura din prima carte;
(6) proba stegulețelor – include itemi sub forma unor steaguri rotite şi reflectate în diferite direcții; sarcina era de a identifica care steaguri sunt identice, printr-o rotire de 90°;
(7) proba identificării numerelor – constă într-o serie de coloane numerice; sarcina subiecților era de a identifica care număr din coloană este identic cu cel din capul coloanei;
(8) proba fețelor – constă în seturi a câte trei fețe umane, dintre care două sunt identice; sarcina avută era de a bifa pe cea care este diferită;
(9) proba citirii în oglindă – fiecare item constă într-un cuvânt scris normal, urmat de altele patru, tipărite ca şi cum ar fi reflectate în oglindă; sarcina era de a identifica care din cele patru cuvinte este identic cu primul.

Bineînțeles că obiectivul principal al aplicării acestor probe era identificarea unor factori comuni din cadrul domeniului inteligenței.

Matricea de corelații dintre cele nouă teste este:

	1	2	3	4	5	6	7	8	9
1. Adunare	1								
2. Înmulțire	.499	1							
3. Serie numere	.394	.436	1						
4. Figuri	.097	.007	.292	1					
5. Cărți	.126	.023	.307	.621	1				
6. Steaguri	.085	.086	.328	.510	.623	1			
7. Identificări număr	.284	.467	.291	.044	.114	.086	1		
8. Fețe	.152	.235	.309	.319	.376	.337	.393	1	
9. Citire în oglindă	.232	.307	.364	.213	.276	.271	.431	.489	1

După alegerea tipului de matrice, orice analiză factorială necesită parcurgerea următorilor paşi:

- ✓ testarea asumpţiilor aferente;
- ✓ alegerea metodei de extragere a factorilor;
- ✓ selectarea numărului de factori extraşi;
- ✓ inspectarea matricei modelului factorial urmată adeseori de rotirea factorilor;
- ✓ interpretarea factorilor.

Cele mai multe aspecte menţionate anterior nu sunt trecute uzual în secţiunea destinată analizei datelor, deşi ele trebuie parcurse pentru a asigura o analiză cât mai corectă. Minimul de informaţii ce trebuie oferit în secţiunea de analiză a datelor unui articol ştiinţific trebuie să atingă următoarele puncte:

- ✓ precizarea într-un paragraf a tipului de analiză întreprinsă (de ex. PCA sau EFA), a metodei de extracţie a factorilor şi a motivaţiei pentru alegerile făcute;
- ✓ precizarea într-un alt paragraf, a numărului de factori extraşi, a criteriilor folosite în acest scop, a tipului de rotire factorială, dacă este cazul, şi a raţiunii alegerii făcute;
- ✓ prezentarea unui tabel sumar care să conţină matricea modelului factorial (pattern matrix) şi gradul de comunalitate a fiecărei variabile. Nicol şi Pexman (1999) oferă un asemenea model, pornind de la standardele Asociaţiei Psihologilor Americani (APA);
- ✓ în cazul rotaţiilor oblice sunt specificate şi corelaţiile între factori.

Pentru acest exemplu, o posibilă formulare a secţiunii de analiză a datelor ar putea fi:

Obiectivul principal al studiului era de a identifica atributele interne care stau la baza celor nouă probe de abilităţi mentale şi influenţează evoluţia acestora. De aceea, utilizând programul SPSS 11.0, am apelat la analiza factorială a axelor principale pentru a extrage factorii comuni.

În urma analizei, atât criteriul scree plot al lui Cattell, cât şi analiza paralelă adaptată după Allen şi Hubard (1986, apud Johnson şi Wickern, 1998) sugerează o soluţie adecvată cu trei factori. Aceştia influenţează împreună 68,48% din totalul dispersiei celor nouă probe. Fiind vorba de factori subsidiari factorului g din domeniul inteligenţei, ne aşteptăm să existe corelaţii între aceştia. În consecinţă, pentru a obţine o structură cât mai simplă, soluţia a fost rotită oblic prin metoda direct oblimin la un delta egal cu zero.

În tabelul de mai jos sunt sintetizate trei categorii de informaţii: gradul de comunalitate a variabilelor, saturaţia acestora în factorii extraşi, cum rezultă din matricea modelului factorial şi corelaţiile dintre factori. Am ales să prezint matricea modelului factorial deoarece este mai simplu de interpretat şi pentru că oferă influenţa fiecărui factor asupra variabilelor în cauză, eliminând medierea celorlalţi factori extraşi. De remarcat însă că aceste saturaţii nu reprezintă coeficienţii de corelaţie simplă, ca în cazul rotaţiilor de tip ortogonal sau a soluţiilor nerotite.

ITEM	FACTORI			Comunalitate
	1	2	3	
1. Adunare	-.06	-.04	.67	.30
2. Înmulţire	.15	.14	.73	.42
3. Serie numere	.08	-.28	.50	.36
4. Figuri	-.01	-.73	.02	.43
5. Cărţi	.05	-.82	.00	.53
6. Steaguri	.05	-.69	.04	.44
7. Identificări număr	.61	.18	.20	.35
8. Feţe umane	.69	-.18	-.12	.36
9. Citire în oglindă	.64	-.06	.03	.35
	Corelaţii între factori			Λ direct oblimin
Factor 1	---			2,18
Factor 2	-.36	---		2,18
Factor 3	.51	-.12	---	1,88

Se observă că primul factor este strâns legat de ultimele trei probe. Toate acestea necesită o examinare rapidă a stimulilor şi identificarea acelora care sunt identici. Din aceste considerente, am putea eticheta acest factor ca viteza de percepţie. Într-un mod similar, se observă că variabilele 4, 5 şi 6 sunt saturate puternic în al doilea factor. Semnul minus este dat de modul invers de cotare a probelor. Aceste trei probe au în comun faptul că necesită abilitatea de a recunoaşte relaţiile dintre diferite forme cu orientare spaţială diferită. Prin urmare, am putea intitula acest factor – relaţii spaţiale. În ceea ce priveşte cel de-al treilea factor, se observă saturarea ridicată în cazul primelor trei variabile şi scăzută pentru celelalte probe. Toate cele trei teste implică calcul numeric. Mai mult, nici una din restul probelor nu necesită calcul aritmetic. În consecinţă, am putea denumi acest factor ca abilităţi de calcul.

Se observă din tabel că cei trei factori nu sunt complet independenţi. Dat fiind vorba de un eşantion de 710 persoane, toate corelaţiile intra factori sunt semnificative la un $p < .01$. Se constată o corelaţie ridicată, în special între viteza de percepţie şi abilităţile de calcul numeric. Aceste corelaţii între factori, vin să sprijine posibilitatea existenţei unui factor supraordonat, fie el factorul g al inteligenţei, aşa cum a fost stipulat de Thurstone.

Elementele prezentate mai sus constituie un minimum necesar de a fi prezentat în secţiunea de analiză cantitativă a datelor. Unii autori susţin şi includerea matricei de corelaţii între variabile. Prin această măsură, cercetătorii interesaţi ar putea reface sau modifica analiza, matricea fiind un element suficient pentru a substitui baza de date. Introducerea matricei de corelaţii este dependentă de spaţiul editorial şi de tipul publicaţiei respective. În ceea ce mă priveşte, recomand prezentarea ei ori de câte ori există spaţiu suficient.

În spatele acestei imagini de „vitrină" se regăsesc o serie întreagă de operaţii, prezentate în detaliu în capitolul de faţă şi care se desfăşoară în următoarea ordine:

(1) alegerea teoretică a tipului de analiză şi a metodei de extragere a factorilor;
(2) verificarea condiţiilor necesare şi inspectarea datelor descriptive;
(3) stabilirea numărului adecvat de factori extraşi şi rotirea factorilor (dacă este cazul);
(4) interpretarea rezultatelor, finalizată cu identificarea naturii factorilor;
(5) sugerarea eventualelor ameliorări necesare în studii viitoare.

5.2. Analiza factorială confirmatorie

Dacă în cazul analizei factoriale exploratorii numărul de factori extraşi se stabilea pe baza unor criterii empirice, analiza factorială confirmatorie apelează la testarea ipotezelor pentru a stabili numărul de factori în care pot fi reduse datele. Bineînţeles că pentru a putea testa aceste ipoteze, cercetătorul trebuie să aibă anumite informaţii legate de numărul de factori care trebuie extraşi, de aceea analiza factorială confirmatorie apare ca o fază ulterioară unui demers descriptiv, de tip exploratoriu.

Problema care va însoţi explicaţiile analizei factoriale confirmatorii priveşte testarea numărul de factori prezenţi în scala BIS-BAS, dezvoltată de Carver şi White (1994), privind sistemele comportamentale activator şi inhibitor. Pentru cei mai puţini familiarizaţi cu aceste constructe, sistemele respective sunt mecanisme neuropsihologice care diferenţiază indivizii în privinţa orientării lor comportamentale predominante, fie spre recompensă şi căutarea plăcerii (BAS), fie spre pedeapsă şi evitarea suferinţei (BIS).

Scala BIS-BAS conţine 20 de itemi şi a fost concepută pentru a surprinde atât sensibilitatea persoanelor la recompensă – BAS, cât şi sensibilitatea persoanelor la pedeapsă – BIS. În urma unei analize factoriale exploratorii, cei doi autori ai probei identifică patru factori primari: BIS (sistemul comportamental inhibitor), BAS-R (sistemul comportamental activator – sensibilitatea la recompensă), BAS-D (sistemul comportamental activator – tendinţe impulsive) şi BAS-FS (sistemul comportamental activator – căutarea plăcerii). Dintre cei patru factori extraşi, ultimii trei corelează între ei, probabil datorită unui factor secundar, denumit BAS. Totuşi, alţi autori au obţinut structuri factoriale diferite. De pildă, Jorm şi colaboratorii (1999), tot în urma unei analize factoriale exploratorii, ajung la concluzia că o soluţie cu doi factori – BIS şi BAS – este la fel de bună ca soluţia cu patru factori găsită de Carver şi White în 1994. Pentru a complica lucrurile, Ross, Millis, Bonebright şi Bailley (2002), în urma unei analize factoriale confirmatorii, găsesc o altă soluţie adecvată. Rezultatele obţinute de aceştia sprijină ideea unei soluţii cu patru factori însă spre deosebire de prima soluţie, toţii factorii sunt independenţi, adică nu corelează între ei. Care dintre soluţii pare cea mai potrivită?

Pentru a răspunde la această întrebare va trebui să aplicăm analiza factorială confirmatorie. Cu ajutorul ei vom testa toate variantele enunţate mai sus şi, chiar mai mult, vom putea testa alte soluţii noi. Stabilirea unei structuri factoriale adecvate este esenţială atât în cercetare, cât şi în psihodiagnostic. Spre exemplu, un individ procedează corect, dacă reuneşte scorurile tuturor itemilor BAS pentru a obţine un rezultat global în cazul în care soluţia corectă implică doi factori BIS şi BAS. Acelaşi lucru poate fi făcut şi în cazul unei soluţii cu patru factori, dintre care cei trei factori BAS corelează între ei. În schimb, cercetătorul va greşi, dacă va calcula un rezultat global BAS în cazul unei soluţii corecte cu patru factori necorelaţi, deoarece între cei trei factori BAS n-ar exista o asociere.

Realizarea unei analize factoriale confirmatorii implică parcurgerea următoarelor etape: (1) specificarea modelelor concurente, (2) identificarea modelelor, (3) estimarea parametrilor statistici şi testarea gradului de potrivire a modelelor, (4) re-specificarea modelelor (Kelloway, 1998).

5.2.1. Specificarea modelelor concurente

Primul pas într-o analiză factorială de acest tip implică specificarea structurilor factoriale ce urmează a fi testate. Ele se aleg pe baza informaţiilor din literatura de specialitate. Spre exemplu, în cazul scalei BIS-BAS vor fi specificate trei asemenea modele: scala BIS-BAS cu doi factori, scala BIS-BAS cu patru factori necorelaţi şi scala BIS-BAS cu patru factori, dintre care cei trei referitori la sistemul comportamental activator sunt corelaţi.

Se obişnuieşte ca cel puţin unul dintre modelele specificate să fie reprezentat grafic într-o formă tipică analizei factoriale confirmatorii. Figura 5.10 conţine o asemenea reprezentare grafică, obţinută prin AMOS (2001), pentru modelul unei scale BIS-BAS cu patru factori, dintre care cei trei, referitori la BAS, corelează între ei. Un alt software, utilizat frecvent pentru a realiza analize factoriale confirmatorii este LISREL (Joreskog şi Sorbom, 1989).

Imaginea poate fi citită uşor. Spre exemplu, se aşteaptă ca primii şapte itemi ai scalei să fie saturaţi în factorul latent BIS. Mai mult, prin absenţa săgeţilor care să lege celelalte variabile latente de primii şapte itemi, se stipulează absenţa saturaţiei itemilor respectivi în cei trei factori BAS. Similar, modelul prezice că itemii 8-12 vor fi saturaţi doar în factorul BAS – recompensă, că itemii 13-16 vor fi atributele manifeste pentru măsurarea factorului BAS – impulsivitate, în timp ce itemii 17-20 vor exprima doar factorul BAS – căutarea plăcerii.

De asemenea, se observă notarea diferită a variabilelor manifeste, adică a datelor observate în realitate, faţă de cea a variabilelor latente, adică a factorilor ce stau în spatele covarianţei itemilor. Astfel, variabilele manifeste sunt reprezentate grafic sub forma unui dreptunghi, rezultând în cazul exemplului dat douăzeci de dreptunghiuri, câte unul pentru fiecare item al testului. În schimb, variabilele latente sunt simbolizate grafic sub forma unui oval, câte unul pentru fiecare factor anticipat. Astfel, un model BIS-BAS cu patru factori, ca cel de faţă, va avea patru

ovale, în schimb un model BIS-BAS cu doi factori va avea doar două asemenea reprezentări.

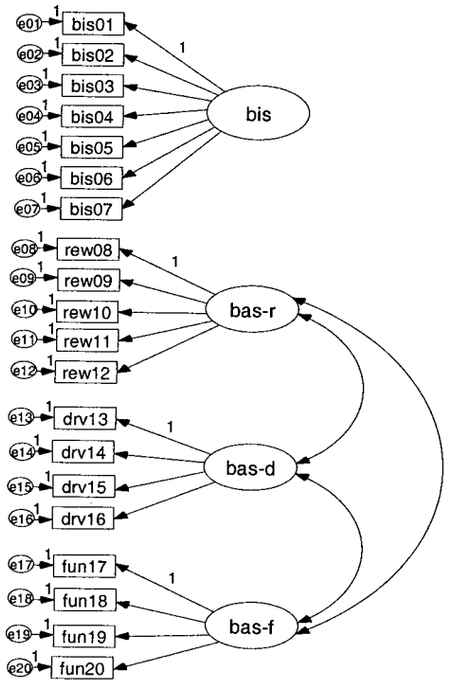


Figura 5.10. *Reprezentarea grafică a modelului BIS-BAS cu patru factori* în AMOS.
Scalele BAS corelează între ele
SPSS este o marcă înregistrată.

O altă observație importantă, pe baza figurii 5.10 este direcția de îndreptare a săgeților. Toate săgețile sunt îndreptate spre itemii probei. De exemplu, în cazul primului item, rezultatul lui BIS01 este în funcție de trăsătura latentă pe care o măsoară itemul respectiv, adică BIS şi de eroarea de măsurare, adică e01. Această eroare poate fi concepută ca dispersie datorată factorilor unici în cazul analizei factoriale.

Corelația dintre doi factori este reprezentată grafic sub forma unei curbe cu săgeți la ambele capete. Spre exemplu, în modelul propus ne aşteptăm să existe o corelație între BAS recompensă (BAS-R) şi BAS impulsivitate (BAS-D) însă stipulăm absența corelației dintre oricare scală BAS şi scala BIS.

În sfârşit, trebuie să distingem între variabile exogene şi variabile endogene. Primele sunt cele care nu depind de nimeni, fiind socotite ca variabile cauză şi se disting grafic prin faptul că nu au nici o săgeată orientată spre ele. În

schimb, variabilele endogene sunt variabile efect, fiind identificate grafic prin faptul că există cel puțin o săgeată orientată spre ele. În exemplul nostru se observă că atât factorii comuni, cât şi erorile de măsurare (factorii unici) sunt variabile exogene, în timp ce itemii probei sunt variabile endogene.

În această etapă sunt specificate toate modelele care urmează a fi testate. Totuşi, din lipsă de spațiu, nu toate modelele beneficiază de reprezentare grafică, deoarece o dată înțelese principiile de reprezentare grafică este suficient să denumim modelul. De pildă, patru factori identici, dar necorelați, implică un desen similar cu cel prezentat în figura 5.10 din care sunt eliminate curbele cu săgeți duble care indicau existența corelațiilor dintre cele trei scale BAS.

5.2.2. Identificarea modelelor

Prin analiza factorială confirmatorie se urmăreşte estimarea saturațiilor itemilor în factorii respectivi, iar pe baza acestora, să se răspundă la întrebarea: în ce măsură saturațiile respective sunt conforme cu modelul specificat? De exemplu, este itemul 5, saturat în factorul BIS şi nesaturat în ceilalți factori, aşa cum era specificat în model? Pentru a estima aceşti parametri, analiza factorială confirmatorie apelează la un set de ecuații structurale (Baer, 2000).

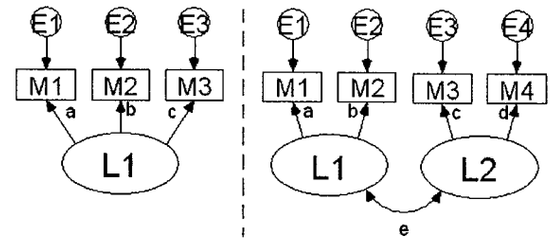


Figura 5.11. *Modalități minimale de reprezentare grafică pentru soluții identificabile*

Prin identificarea modelelor se înțelege în ce măsură acestea, prin structura lor, pot conduce la soluții identificabile. Fiind vorba de ecuații structurale, se pune problema matematică dacă există: (1) mai multe necunoscute decât ecuații, (2) atâtea ecuații câte necunoscute, (3) mai multe ecuații decât necunoscute. Vom discuta puțin mai târziu implicațiile acestor situații distincte. Deocamdată vom clarifica problema identificării.

De exemplu, în imaginea din stânga trebuie estimată saturația celor trei variabile M1, M2 şi M3 în factorul L1, adică trebuie identificate valorile a, b şi c. Pentru a le identifica avem la dispoziție trei ecuații:

$$M1 = aL1 + e1$$
$$M2 = bL1 + e2$$
$$M3 = cL1 + e3$$

Algebric aceste ecuații pot fi transpuse în matricea de covarianță:

$$\begin{bmatrix} \operatorname{cov} M1M1 & \operatorname{cov} M2M1 & \operatorname{cov} M3M1 \\ \operatorname{cov} M1M2 & \operatorname{cov} M2M2 & \operatorname{cov} M3M2 \\ \operatorname{cov} M1M3 & \operatorname{cov} M2M3 & \operatorname{cov} M3M3 \end{bmatrix}$$

unde pe diagonala principală se regăsesc, de fapt, dispersiile variabilelor M1, M2 şi M3 (covarianța dintre o variabilă şi ea însăşi).

Am insistat asupra acestor aspecte tehnice pentru a facilita înțelegerea identificării unui model. Astfel, pentru ca un model să fie identificabil, adică testabil, trebuie ca numărul elementelor non-redundante din matricea de covarianță să fie mai mare decât suma dintre totalul variabilelor exogene şi totalul parametrilor (saturațiilor) de estimat. Regula este cunoscută sub numele de criteriul lui t (Baer, 2000).

În matricea de mai sus, corespunzătoare imaginii din stânga a figurii 5.11, există şase elemente non-redundante (covariațele M1M2, M1M3 şi M2M3 şi dispersiile lui M1, M2 şi M3). De asemenea, putem distinge din imagine prezența a trei variabile endogene (itemii M1, M2 şi M3), patru variabile exogene (L1, e1, e2 şi e3) şi respectiv 3 parametri de estimat (a, b şi c). Suma dintre ultimele două categorii indică 7 elemente. Totuşi, în analiza factorială confirmatorie, dispersia lui L1 este ținută constantă la 1, pentru a standardiza saturările obținute, de aceea suma devine şase şi indică o egalitate cu numărul elementelor non-redundante. O altă posibilitate de acțiune nu implică prestabilirea dispersiei variabilei latente la 1, ci presupune calcularea ei. În schimb, se alege o valoare de referință între coeficienții de saturație ce urmează a fi estimați, căreia i se atribuie o valoare constantă, de obicei egală cu 1. Astfel, numărul parametrilor de estimat rămâne acelaşi. Ambele situații sunt denumite „abia identificabile" (în engleză „just-identified") şi implică un număr de ecuații egal cu numărul de necunoscute. Datorită acestui fapt, soluțiile unice rezultate în asemenea situații vor indica un model perfect, ce nu va putea fi comparat cu altele concurente. Situația prezintă, din punctul de vedere al numărului de elemente, minimul necesar, dar nu şi suficient pentru a aplica analiza factorială confirmatorie.

În altă ordine de idei, o variabilă latentă cu doar două variabile manifeste implică trei elemente non-redundante (covarianța M1M2, dispersia M1 şi dispersia M2) şi patru elemente de estimat (coeficienții de saturație a şi b ai variabilelor M1 şi M2 în L1, respectiv factorii unici e1 şi e2). O asemenea situație corespunde unui număr de ecuații mai mic decât numărul de necunoscute; prin urmare, modelul nu este testabil, deoarece soluțiile nu pot fi identificate. Totuşi, o astfel de situație poate fi reparată, fie prin impunerea unor constricții de egalitate precum a = b, fie prin includerea altei variabile latente suplimentare care corelează cu prima. Imaginea din dreapta figurii 5.11. corespunde ultimei situații, deoarece cele 9 elemente de estimat (coeficienții a, b, c, d şi e, respectiv erorile e1, e2, e3 şi e4)

pot fi identificate prin zece covariațe non-redundante (covariațele M1M2, M1M3, M1M4, M2M3, M2M4, M3M4 şi dispersiile M1, M2, M3 şi M4). Această situație este dezirabilă, ca oricare alta în care numărul de ecuații este mai mare decât numărul de necunoscute, oferind posibilitatea testării şi comparării mai multor modele concurențiale.

De exemplu, în cazul modelului factorial al scalei BIS-BAS prezentat în figura 5.10 există 210 elemente non-redundante, dintre care 20 de dispersii corespunzătoare celor 20 de itemi şi 190 de covariațe (obținute pe baza formulei k(k-1)/2, unde k reprezintă numărul de itemi ai scalei). De asemenea, există 43 de elemente de estimat (20 de saturații factoriale, trei corelații între factorii de tip BAS şi 20 de erori sau factori unici). Se poate observa că soluția poate fi identificată, deoarece numărul elementelor non-redundante este mai mare decât numărul elementelor de estimat. Diferența de 167 dintre ele reprezintă, de fapt, numărul de grade de libertate pentru testarea modelului respectiv.

Înainte de a trece la discutarea soluției găsite şi de a o compara pe aceasta cu modelele concurente, trebuie să distingem faptul că două modele pot fi în relație de tip cuib sau nu.

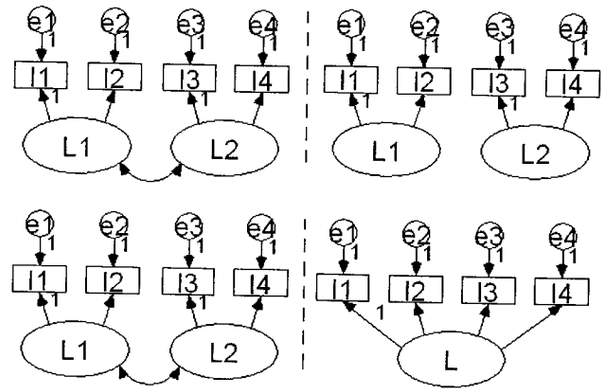


Figura 5.12. *Exemple de modele factoriale aflate în relație de tip cuib*

Dacă privim cele două modele despărțite prin linie punctată din partea de sus a figurii 5.12, vom vedea că cel de-al doilea model este identic cu primul, lipsindu-i doar simbolul prezenței unei corelații între cele două variabile latente L1 şi L2. O asemenea relație între modele se numeşte de tip cuib (în engleză „nested"), deoarece putem spune că modelul al doilea este inclus în primul. De fapt, vom obține două modele identice, dacă vom constrânge pe cel de-al doilea să aibă o corelație egală cu 0.

Deşi mai puțin intuitiv, şi între cele două modele din partea de jos a figurii 5.12 există o relație de tip cuib. Primul model poate fi transformat matematic în cel de-al doilea, dacă vom constrânge cele două variabile latente L1 şi L2 să aibă o corelație egală cu 1. În schimb, două modele nu se află în relație de tip cuib dacă nu putem reduce unul din modele la celălalt.

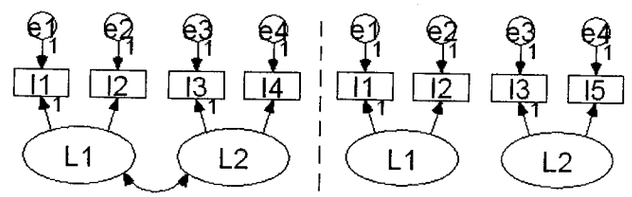


Figura 5.13. *Exemple de modele factoriale care nu se află în relație de tip cuib*

Cele două modele nu se află în relație de tip cuib deoarece variabila latentă L2 este măsurată prin itemii 3 şi 4 (în stânga), în timp ce în dreapta, L2 este măsurat de itemii 3 şi 5. Astfel, nici unul dintre modele nu este reductibil la celălalt.

Distincția dintre modelele aflate în relație de tip cuib şi cele care nu se regăsesc într-o asemenea relație este importantă, deoarece primul tip permite compararea directă a modelelor din perspectiva unui test de semnificație.

5.2.3. Stabilirea gradului de adecvare a modelelor

Ca în orice analiză factorială, matricea de covarianță sau cea de corelație constituie inputul demersului statistic. Un model factorial este considerat bun, dacă prin parametrii estimați reuşeşte să refacă cât mai bine matricea datelor inițiale.

Pentru a testa gradul de adecvare a unei soluții factoriale există o multitudine de indicatori statistici, dintre care vom prezenta câțiva în continuare. Ei pot fi împărțiți în două categorii mari: indici absoluți şi indici comparativi. Cei din urmă pot fi de asemenea împărțiți în indicatori comparativi între modele aflate în relație de tip cuib şi indicatori de comparare a modelelor care nu se află în relații de tip cuib. În sfârşit, acestea din urmă pot fi divizate tot în două categorii: indicatori ai gradului de potrivire şi indicatori ai gradului de parcimonie a modelului.

5.2.3.1. Indicatori absoluți

Aceşti indicatori urmăresc în ce măsură un anumit model poate reproduce matricea de covarianță (corelație). Primul indicator este testul χ^2 al gradului de potrivire. O valoare nesemnificativă a acestuia indică absența unor diferențe semnificative între matricea datelor brute şi matricea obținută pe baza legăturilor specificate în model, de această dată fiind dezirabilă situația unui χ^2 nesemnificativ

statistic. Din păcate, valoarea lui χ^2 este afectată puternic de volumul eşantionului, astfel că există mari şanse să obținem un χ^2 semnificativ în cazul testării unui număr mare de subiecți, cum se întâmplă frecvent în studiile ce implică analiza factorială. Totuşi, χ^2 este deosebit de util, deoarece stă la baza mai multor indicatori absoluți care depăşesc această limită. Utilizând fişierul „analiza factoriala confirmatorie.sav" în programul statistic AMOS 4.0, vom obține în cazul scalei BIS-BAS cu patru factori (factorii BAS corelează între ei) o valoare a discrepanței $\chi^2(167) = 375.85$, $p < .001$ însă rezultatul trebuie interpretat cu precauție datorită numărului mare de date observate (N=345).

Un alt indicator absolut este RMR (în engleză „root mean squared residuals"). Conform lui Steiger (apud Kelloway, 1998) valorile mai mici de .10 indică un model bun, iar valorile sub .05 un model foarte bun. Printr-un model bun se înțelege că modelul respectiv reproduce (explică) bine evoluția datelor observate. Pentru scala BIS-BAS, RMR a fost de .04.

Tot un indicator rezidual absolut este şi RMSEA (în engleză „root mean squared error of aproximation) care are aceleaşi proprietăți şi valori referențiale. În plus, acesta are avantajul că oferă un interval de încredere pentru valoarea calculată. Dacă limita maximă a intervalului nu depăşeşte .08, avem de-a face cu un model destul de bun. În problema noastră, modelul BIS-BAS cu patru factori, dintre care cei trei legați de BAS corelează între ei, RMSEA a fost .06, iar intervalul de încredere de 90% este cuprins între .052 şi .068, valori ce indică, din nou, o bună potrivire a modelului.

Un alt indicator absolut, probabil cel mai cunoscut, este GFI (în engleză „index of goodness of fit"). Acesta ia valori cuprinse între 0 şi 1 însă sunt dorite valori peste .90 pentru a considera că modelul testat este unul bun. Totuşi, după unele criterii mai liberale şi valorile situate peste .85 pot semnala un model factorial bun. GFI este un indicator sensibil, fiind dependent de numărul de date colectate şi de complexitatea modelului. De aceea, se poate calcula şi un indicator ajustat al acestuia, simbolizat AGFI, care are aceleaşi valori de referință. În cazul modelului BIS-BAS testat de noi, am obținut un GFI de .90 şi un AGFI de .88, valori care indică un model destul de bun.

Merită să subliniem faptul că aceşti indicatori se calculează pentru fiecare model concurent anticipat. De obicei aceste valori se trec într-un singur tabel, pentru a facilita analiza gradului de potrivire a fiecărui model față de datele observate.

Tabelul 5.6. Valorile principalilor indicatori absoluți pentru cele trei modele testate

Modelul testat	χ^2	df	p	RMR	RMSEA	GFI	AGFI
Doi factori necorelați	915.3	170	.001	.07	.113 [.106 - .120]	.76	.70
Patru factori necorelați	431.7	170	.001	.06	.067 [.059 - .075]	.89	.86
Patru factori (scalele BAS corelează)	375.8	167	.001	.04	.062 [.052 - .068]	.90	.88

Din tabel se poate observa că dintre cele trei modele propuse spre a fi testate, doar ultimele două par a fi adecvate pentru a descrie datele culese. Primul model, al conceperii scalei BIS-BAS ca o probă ce măsoară doi factori primari este inadecvat, valorile indicatorilor absoluți fiind departe de valorile de referință amintite mai sus. Între cele două modele rămase în cursă sunt diferențe mici, observându-se un avantaj uşor în favoarea ultimului model. Pentru a compara aceste modele, putem să apelăm la indicatorii de comparare.

5.2.3.2. Indicatori de comparare

Indicatori de comparare a modelelor aflate în relație de tip cuib

Indicatorul de bază în astfel de cazuri este χ^2 al diferențelor sau Δ(delta) χ^2. Aşa cum îi spune numele, el se calculează făcând diferența între χ^2, respectiv df corespunzători modelelor aflate în relație de tip cuib. Pe baza informațiilor prezentate anterior distingem două perechi de modele „cuib": unu cu trei, respectiv doi cu trei. Cum primul model s-a dovedit a fi ineficient, vom compara doar modelele doi şi trei. Vom obține $\Delta\chi^2$ (3) = 55.9, p < .001, ceea ce indic• diferențe semnificative între cele două modele, mai bun fiind cel de-al treilea model, adică cel care are un χ^2 mai mic. Pe baza acestor rezultate putem concluziona c•, deşi ambele modele se dovedesc eficiente în a reproduce datele observate, modelul al treilea este semnificativ mai bun decât modelul cu patru factori necorelați.

Indicatori de comparare a modelelor în general

Când modelele nu se află în relații de tip cuib, nu putem aplica χ^2 al diferențelor. În astfel de cazuri putem aplica alți indicatori. De fapt, aceşti indicatori pot fi aplicați indiferent de tipul relației dintre două modele, însă în cazul relațiilor de tip cuib se preferă utilizarea lui χ^2 al diferențelor, datorită testului de semnificație asociat cu acesta.

Dacă indicatorii absoluți testau diferențele dintre modelul stipulat şi un model saturat (întotdeauna adevărat ca în situația unui model abia-identificabil), indicatorii de comparație testează diferențele dintre modelul testat şi un model nul, slab. Există mai mulți indicatori utilizați în acest sens precum NFI şi NNFI (în engleză „normed fit index", respectiv „non-normed fit index") dezvoltați de Bentler şi Bonnet, CFI (în engleză „comparative fit index") dezvoltat de Bentler, IFI (în engleză „incremental fit index") al lui Bollen sau RFI (în engleză „relative fit index") al lui Marsh (Kelloway, 1998). Toți indicatorii, cu excepția NNFI sunt cuprinşi între 0 şi 1, valorile peste .90 ori .85 fiind dezirabile pentru a considera un model bun.

Tabelul 5.7. Valorile principalilor indicatori de comparație pentru cele trei modele testate

Modelul testat	NFI	RFI	IFI	CFI
Doi factori necorelați	.48	.42	.53	.53
Patru factori necorelați	.75	.73	.84	.83
Patru factori (scalele BAS corelează)	.80	.77	.87	.87

Se observă, şi în acest caz, diferența valorică dintre primul model şi celelalte două. Totuşi, valorile indicatorilor de comparare sunt mai mici decât cele referențiale, ceea ce aruncă o umbră de îndoială asupra celor trei modele testate. Dintre ele, ultimul model se apropie cel mai mult de valorile dezirabile, acestea fiind atinse chiar în cazul IFI şi CFI şi demonstrând încă o dată superioritatea acestui model în comparație cu celelalte.

Indicatori de comparare a parcimoniei modelelor

Cu siguranță că un model, cu cât este mai complex, cu atât reproduce mai bine realitatea. Transpus în termeni de indicatori de comparare, ideea implică o favorizare a modelelor complexe în dauna celor simple, primele având şanse mai mari de a obține indicatori de comparare dezirabili. În ştiință însă sunt favorizate explicațiile mai simple, dorindu-se eliminarea factorilor redundanți. De aceea, pentru a elimina distorsiunile indicatorilor de comparație aflați sub influența complexității modelului, se pot calcula alți indicatori meniți să avantajeze modelele mai simple, în cazul în care cele complexe nu oferă un avantaj semnificativ.

Mulți dintre indicatorii absoluți sau de comparație au formule ajustate pentru a avantaja soluțiile cu un grad ridicat de parcimonie. Astfel, între două soluții la fel de bune, cele mai simple vor fi avantajate din perspectiva acestor indicatori. Cei mai întâlniți sunt PNFI (în engleză „parsimony adjusted NFI") şi PCFI („parsimony adjusted CFI), aceştia fiind doi indicatori de bază oferiți de AMOS. Ambii au valori cuprinse între 0 şi 1 însă spre deosebire de indicatorii neajustați nu au valori de referință prestabilite. De aceea, ca strategie de utilizare se obişnuieşte avantajarea acelor modele care au o valoare mai ridicată. Acelaşi rol îl joacă alți doi coeficienți AIC („Akaike Information Criterion") şi CAIC („Consistent Akaike Information Criterion"). În cazul acestora este dezirabil modelul cu valori mai mici, adică modelul mai simplu. Se observă şi din tabelul 5.8 că ultimele două soluții sunt mai bune, iar dintre ele, se constată o uşoară inferioritate a modelului cu patru factori necorelați.

Tabelul 5.8. Valorile principalilor indicatori de comparație a simplității modelelor testate

Modelul testat	PNFI	PCFI	AIC	CAIC
Doi factori necorelați	.43	.47	995.4	1189.4
Patru factori necorelați	.68	.75	511.7	705.4
Patru factori (scalele BAS corelează)	.69	.76	431.7	630.3

Pe baza tuturor acestor indicatori putem concluziona că modelul factorial adecvat pentru a descrie scala BIS-BAS pe eşantionul de 345 studenți români este cel de-al treilea, reprezentat grafic în figura 5.10.

Prezentarea rezultatelor se poate face, fie sub formă grafică, fie sub formă tabelară. Deosebirea dintre analiza factorială exploratorie şi cea confirmatorie este că în ultimul caz se calculează saturările doar în factorii anticipați. De exemplu, se

aşteaptă ca itemul 1 să fie saturat în factorul BIS şi nesaturat în ceilalți factori. Prin urmare, se va estima doar saturarea în BIS, celelalte saturații ale itemului în factorii BAS fiind prestabilite la zero. Acelaşi lucru se întâmplă şi în privința corelațiilor dintre factori. Deoarece modelul presupune absența corelației dintre scala BIS şi oricare din scalele BAS, corelațiile dintre acestea sunt prestabilite la zero. În schimb, sunt estimate corelațiile dintre cele trei scale de tip BAS, deoarece acelaşi model anticipa corelații semnificative între cei trei factori.

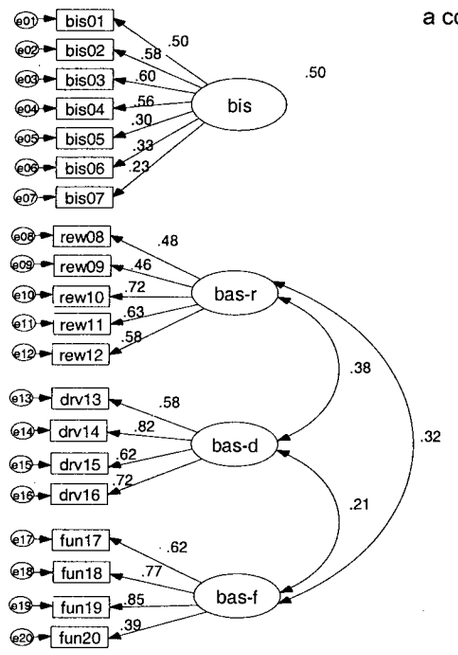


Figura 5.14. *Structura factorială BIS-BAS*

Tabelul 5.9. Matricea modelului factorial BIS-BAS şi a corelațiilor inter-factori

Item	BIS	BAS-R	BAS-D	BAS-F
01	.50			
02	.58			
03	.60			
04	.56			
05	.30			
06	.33			
07	.23			
08		.48		
09		.46		
10		.72		
11		.63		
12		.58		
13			.75	
14			.82	
15			.62	
16			.72	
17				.62
18				.77
19				.85
20				.39
BAS-R	--			
BAS-D	.38	--		
BAS-F	.32	.21	--	

5.2.4. Modificarea modelului

La fel ca în cazul analizei factoriale exploratorii, şi în cazul celei confirmatorii problema nu se încheie o dată cu testarea modelelor şi oferirea rezultatelor modelului selectat. Demersul poate fi continuat cu analiza eventualelor modificări în modelul selectat. Ce parametri noi pot fi estimați? La ce parametri deja estimați se poate renunța? La asemenea întrebări se poate răspunde apelând la indicatorii modificării.

Aceşti indicatori Langrange calculează gradul de modificare a lui χ^2 al modelului o dată cu introducerea sau eliminarea fiecărui parametru posibil. Dintre toate valorile calculate, vor trezi interesul cele care îndeplinesc două condiții:

conduc la o scădere substanțială a lui χ^2 (1) şi cele care au o semnificație teoretică importantă (2). Dacă ultima nu ridică probleme, fiind variabilă de la caz la caz, prima este mai ambiguă. Aşa cum remarca Kelloway (1998), nu se poate vorbi de o valoare standard, de exemplu un MI de 5 nu înseamnă mare lucru la un χ^2 de 300, însă modificarea devine importantă în cazul unui χ^2 de 100. Din experiență, merită a fi analizate propunerile de modificare a modelului care au indicatorii modificării mai mari decât raportul $\chi^2/25$. În cazul modelului nostru, $\chi^2(167) = 345.8$, $p < .001$; rezultă o valoare de referință a MI de 13,83.

Covariances:			M.I.	Par Change	Regression Weights:			M.I.	Par Change
e15	<-->	e16	23,011	0,105	bis15	<--	bis17	14,609	0,148
e13	<-->	bas-f	17,644	-0,070	bis13	<--	bis19	15,390	-0,122
e13	<-->	e14	14,304	0,048	bis13	<--	bis17	15,572	-0,112
e08	<-->	e09	16,105	0,036	bis01	<--	bis08	13,992	0,334
Variances:			M.I.	Par Change					

Figura 5.15. *Output-ul oferit de AMOS cu valorile indicilor modificării mai mari de 13,83*
SPSS este o marcă înregistrată.

Dintre cele opt propuneri de modificare observăm că nu există nici una care să aibă o importanță teoretică, majoritatea aspectelor privind corelații între erori. Singura modificare care ar fi putut avea loc ar fi fost eliminarea itemului 7 din scală, datorită unei saturații mai reduse în BIS, sub valoarea recomandată de .30, dar indicatorul de modificare în urma renunțării la această relație este insignifiant, fiind chiar mai mic decât 3. O asemenea valoare nu conduce decât la modificări minuscule în evoluția principalilor indicatori, aşa cum reiese din tabelul 5.10.

Tabelul 5.10. Evoluția comparativă a principalilor indicatori de testare a modelelor

Modelul testat	χ^2	df	GFI	AGFI	RMSEA	NFI	PNFI
BIS-BAS patru factori – 19 itemi	349.7	149	.90	.88	.063 [.054 - .071]	.80	.69
BIS-BAS patru factori – 20 de itemi	345.8	167	.90	.88	.062 [.052 - .068]	.80	.69

Această situație recomandă păstrarea itemului 7 în test, deoarece eliminarea sa nu conduce la o îmbunătățire a modelului. În urma renunțării la itemul respectiv nu se constată un avantaj comparativ al acestui model în dauna

celui cu 20 de itemi. De aceea se recomandă păstrarea variantei de 20 de itemi a scalei BIS-BAS.

De menționat, totuşi, că şi în cazul în care acceptăm modificările pe baza acestor indicatori, ele nu au un caracter confirmativ, ci doar exploratoriu.

5.2.5. Aspecte metodologice ale aplicării analizei factoriale confirmatorii

În finalul capitolului vom prezenta pe scurt informații legate de volumul eşantionului în analiza factorială confirmatorie, de tipul de date colectate ce pot fi analizate factorial, precum şi de alte condiții ale distribuției datelor ce trebuie îndeplinite pentru a putea aplica această tehnică.

În ce priveşte numărul de subiecți necesari pentru a aplica această analiză se păstrează regulile din analiza factorială exploratorie. Astfel, parametrii nu pot fi estimați adecvat sub 100 de subiecți şi/ sau în condițiile în care nu există un raport de cinci-zece subiecți la o variabilă. Totuşi, ar fi de dorit ca numărul de subiecți testați să fie de cel puțin 200 de persoane pentru modele factoriale de complexitate medie.

Datele care pot fi analizate prin această metodă pot fi de tip numeric sau ordinal. Astfel, răspunsurile nominale dihotomice nu sunt cele mai potrivite pentru o astfel de analiză, deşi printr-un artificiu determinat de proprietățile datelor dihotomice ar putea fi tratate ca valori numerice binare 1 sau 0. Totuşi, în acest caz ar apărea o altă problemă ce ține de imposibilitatea de a obține o distribuție normală multivariată a datelor dihotomice, condiția respectivă fiind necesară pentru a aplica metoda estimării verosimilității maxime, cea mai întâlnită procedură în analiza factorială confirmatorie. Rămâne ca o alternativă posibilitatea grupării itemilor în pachete de itemi (în engleză „parcels") care să fie analizate factorial.

După Bryant şi Yarnold (1995) datele cu amplitudine scăzută sau distribuții asimetrice conduc la atenuarea saturației itemilor în factorii respectivi. În cazul unor distribuții multivariate foarte anormale, atât metoda estimării verosimilității maxime (în engleză MLE - „maximum likelihood estimation"), cât şi metoda celor mai mici pătrate generalizate (în engleză GLS – „generalized least squared") vor produce indicatori absoluți şi comparativi distorsionați, iar aplicarea altor metode cum ar fi ADF (în engleză „assumption distribution free") necesită un număr enorm de subiecți, de câteva mii.

În sfârşit, nu toate elementele discutate mai sus trebuie prezentate în textul articolelor trimise spre publicare, datorită spațiului redus. În general, ca în orice problemă incluzând ecuații structurale latente, informația minimă oferită trebuie să cuprindă:

1. o reprezentare grafică a unui model factorial structural, de obicei a celui rezultat din datele autorilor probei respective; această reprezentare poate fi eliminată din motive practice în cazul probelor lungi, peste 50-100 de itemi; în acea situație, figura poate fi înlocuită cu un tabel care să prezinte atât

factorii aşteptați, cât şi numărul de identificare al itemilor saturați în aceştia; totuşi, graficul poate fi prezentat în cazul în care unitatea de bază nu este itemul, ci pachetul de itemi similari;
2. informații sumare despre tipul de matrice utilizată, despre metoda folosită pentru estimarea parametrilor şi despre tratamentul datelor lipsă şi a celor neobişnuite;
3. prezentarea într-un tabel a principalilor indicatori absoluți şi comparativi pentru fiecare model testat, la care vom adăuga parametrii modelului selectat modificat, dacă este cazul; din pleiada de indicatori menționați, pe baza proprietăților lor sau a frecvenței de utilizare, recomand prezentarea următorilor indici: χ^2(df) cu pragul de semnificație asociat, GFI, AGFI, intervalul de încredere RMSEA, NFI, CFI, PNFI şi PCFI;
4. compararea, dacă este cazul, a soluțiilor aflate în relații de tip cuib;
5. ultima piesă necesară implică prezentarea soluției alese din perspectiva parametrilor standardizați estimați implicând gradul de saturare a itemilor în factorii respectivi şi, dacă este cazul, a nivelului de corelație între factori (vezi figura 5.14.).

BIBLIOGRAFIE

Baer, D. (2000). *Latent variable structural equation models.* A student's handbook at ICPSR Summer Program, Ann Arbor, USA.

Browne, M.W., Cudeck, R., Tateneni, K., Mels, G. (1998). *CEFA: Comprehensive exploratory factor analysis.* Găsit la: http://quantrm2.psy.ohio-state.edu/browne/

Bryant, F.B., Yarnold, P.R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. În L.G. Grimm şi P.R. Yarnold (eds.) *Reading and understanding multivariate statistics* (p.99-136). Washington DC: American Psychological Association.

Carver, C.S., White, T.L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, 67, 319-333.

Costa, P.T., McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI). Professional Manual.* Odessa, Florida: Psychological Assessment Resources, Inc.

Field, A. (2000). *Discovering Statistics using SPSS for Windows.* London: Sage Publications.

Finch, J.F., West, S.G. (1997). The investigation of personality structure: statistical models. *Journal of Research in Personality*, 31, 439-485.

Garson (2000). *Factor analysis.* Găsit la: http://www2.chass.ncsu.edu/garson/pa765/logistic.htm

Hair, J.F. Jr., Anderson, R.E., Tatham, R.L., Black, W.C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Iacobucci, D. (1994). Classic Factor Analysis. În R.P. Bagozzi (ed.) *Principles of marketing research.* Cambridge, MA: Blackwell.

Johnson, R.A., Wickern, D.W. (1998). *Applied multivariate statistical analysis (4th ed.).* Englewood Cliffs, NJ: Prantice-Hall.

Joreskog, K.G., Sorbom, D. (1989). *LISREL-7 user's reference guide.* Mooresville, IN: Scientific Software.

Jorm, A.F., Chriestensen, H., Henderson, A.S., Jacomb, P.A., Korten, A.E., Rodgers, B. (1999). Using the BIS/BAS scales to measure behavioral inhibition and behavioral activation: Factor structure, validity, and norms in a large community sample. *Personality and Individual Differences,* 26, 49-58.

Kachigan, S.K. (1991). *Multivariate statistical analysis. A conceptual introduction* (2nd ed.). New York: Radius Press.

Kelloway, E.K. (1998). *Using LISREL for structural equation modeling. A researcher's guide.* Thousand Oaks, USA: Sage Publications.

Kline, P. (1993). *The handbook of psychological testing.* London: Routledge.

Mărgineanu, N. (1938). *Analiza factorilor psihici.* Cluj-Napoca: Editura Institutului de Psihologie a Universității din Cluj.

Meulman, J.J., Heiser, W.J. (2001). *SPSS Categories 11.0.* Chicago: SPSS Inc.

Miclea, M. (1994). *Psihologie cognitivă.* Cluj-Napoca: Casa de Editură Gloria.

Nicol, A.A.M., Pexman, P.M. (1999).*Presenting your findings. A practical guide for creating tables.* Washington DC: American Psychological Association.

Pedhazur, E.J, Schmelkin, L. (1991). *Measurement, design, and analysis. An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Poenaru, R. (1980). Didactogenia în şcoală. Pentru prevenirea şi combaterea didactogeniilor din învățământ. *Revista de pedagogie (9),* 50-51.

Poenaru, R, Sava, F. (1998). *Didactogenia în şcoală. Aspecte deontologice, psihologice şi pedagogice.* Bucureşti: Editura Danubius.

Ross, S.R., Millis, S.R., Bonebright, T.L., Bailley, S.E. (2002). Confirmatory factor analysis of the Behavioral Inhibition and Activation Scales. *Personality and Individual Differences,* 33, 861-865.

Sava, F. (2002). Causes and effects of teacher conflict-inducing attitudes towards pupils: a path analysis model. *Teaching and teacher education, 17*(8), 1007-1021.

Stills, D.L. (Ed.).(1989). *Interntional encyclopedia of the social sciences: Biographical supplement* (Vol.18). New York: Macmillan.

Tucker, L.R., MacCallum, R.C. (1997). *Exploratory factor analysis. An online book.* Găsit la: http://quantrm2.psy.ohio-state.edu/maccallum/factornew.htm

Wegener, D.T., Fabrigar, L.R. (2000). Analyzing nonexperimental data. În H.T. Reis şi C.M. Judd (Eds.). *Handbook of research methods in social and personality psychology* (p.412-450). Cambridge: Cambridge University Press.

Young, P. (2002). ViSta. Găsit la: http://forrest.psych.unc.edu/research/

Zwick, W.R., Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin,* 99(3), 432-442.

CAPITOLUL 6

REGRESIA LINIARĂ

Predilecția firii umane de a consulta horoscopul sau de a asculta opiniile avizate ale unui „expert în ghicit viitorul într-o ceaşcă de cafea" s-a menținut în ciuda dezvoltării tehnologice şi ştiințifice. Regresia statistică, deşi nu la fel de populară ca metodele amintite anterior, îşi propune, de multe ori, acelaşi obiectiv: o predicție cât mai exactă a performanțelor din viitor pe baza unor indicatori din prezent sau trecut. Denumirea de regresie poate induce în eroare specialiştii în psihologie, obişnuiți să utilizeze acest termen mai ales din psihanaliză, unde prin regresie se înțelegea o revenire spre trecut. În statistică, prin regresie ne direcționăm frecvent în sens opus, spre viitor, încercând să prevedem cât mai exact evoluția viitoare a lucrurilor. De exemplu, dacă suntem un psiholog organizațional, va trebui să prezicem, în procesul de selecție a personalului, care persoană merită să fie angajată dintr-un grup de candidați, pe baza unei estimări a performanței viitoare a acestora. Similar, dacă suntem un psiholog şcolar şi suntem nevoiți să identificăm un grup cu risc crescut de a consuma droguri, va trebui să estimăm, pe baza caracteristicilor lor actuale, evoluția lor ulterioară. Ambele situații pot fi rezolvate prin apelul la tehnica statistică denumită regresie, în cazurile respective fiind vorba de o ***regresie cu scop predictiv***. Capacitatea de predicție constituie de altfel una dintre caracteristicile definitorii ale unei discipline pentru a fi considerată o ştiință.

Un alt tip de regresie, similară în multe privințe cu cea utilizată în scop predictiv, este ***regresia în scop explicativ***. Din titulatură reiese diferența esențială dintre cele două scopuri majore ale regresiei. Prima este utilizată pentru a prezice performanțele viitoare ale unui grup de oameni pe baza unuia sau a mai multor indicatori actuali, denumiți predictori. Cea de-a doua, vizează explicarea influenței unui set de factori asupra unei variabile criteriu, cunoscându-se existența unei relații cauzale între acestea. Un exemplu tipic de utilizare a regresiei în scop explicativ este acela al identificării influenței fumatului în perioada prenatală asupra dezvoltării emoționale a copilului în primii ani de viață. Există o multitudine de factori care influențează dezvoltarea emoțională a copilului. Regresia explicativă nu face altceva decât să izoleze (controleze) statistic influența factorilor respectivi, pentru a vedea în ce măsură fumatul contribuie în plus la acest lucru. De pildă, dacă se cunoaşte că tipul de ataşament, stilul educațional al părinților, prezența sau absența altor frați sau surori etc. influențează dezvoltarea emoțională a copilului, prin regresie de tip explicativ se pot controla toate aceste influențe pentru a se observa dacă fumatul mamei în perioada gravidității contribuie, în plus, la dezvoltarea emoțională normală a copilului.

Diferențieri între cele două tipuri de regresie, în scop predictiv, respectiv în scop explicativ, apar şi la nivel terminologic. În cazul ultimei, se pot folosi termenii de **variabilă independentă** şi **variabilă dependentă**, iar în cazul primului tip, se folosesc sintagmele **predictor** şi **criteriu**. Această diferențiere terminologică are la bază logica metodologică folosită în cazul celor două tipuri de regresie. Astfel, regresia în scop explicativ urmăreşte să vadă dacă un anumit factor influențează evoluția variabilei dependente, prin controlarea influenței exercitate de alte variabile incluse în model. Dacă prin controlarea acestor influențe, variabila (variabilele) independentă(e) de interes contribuie semnificativ la evoluția criteriului, am putea deduce existența unei relații între acestea. Dacă relația este susținută de o teorie relevantă în domeniu sau de o logică temporală sau spațială a relațiilor dintre ele, ne vom permite să utilizăm termenii de variabilă independentă şi variabilă dependentă, chiar dacă nu avem de a face cu un design experimental. Prin urmare, întrebarea principală la care încearcă să răspundă regresia utilizată în scop explicativ este: „Contribuie acest factor la evoluția variabilei dependente, în condițiile în care am controlat statistic influența altor factori asupra acesteia?".

Nu acelaşi lucru se face şi în cazul utilizării regresiei în scop predictiv. Cu toate că nu este exclusă posibilitatea existenței unei relații cauzale între predictori şi criteriu, nu acest lucru constituie interesul principal al analizei de regresie. Aşa cum îi spune şi numele, utilizarea regresiei în scop predictiv urmăreşte găsirea unei ecuații care să estimeze cel mai bine evoluția criteriului. Concret, dacă criteriul urmărit este nivelul de depresie, pot fi incluşi între predictori tot felul de indicatori care nu se află în relație cauzală cu criteriul, dar estimează bine nivelul acestuia. De exemplu, includerea preferinței pentru culoarea neagră, scăderea apetitului sexual etc., nu sunt cauze ale depresiei, dar împreună cu alți asemenea indicatori pot fi utilizați eficient în a prezice nivelul de depresie al unei persoane. Prin urmare, întrebarea principală la care încearcă să răspundă regresia utilizată în scop predictiv este: „Care sunt predictorii care ne ajută să estimăm cel mai bine evoluția criteriului?". În cazuri rare, putem vorbi şi de o regresie cu scop mixt, dacă obiectivele cercetării vizează atât aspecte estimative, cât şi explicative.

O altă diferențiere terminologică importantă este între **regresie liniară simplă** şi **regresie multiliniară**. În primul caz, vorbim de o regresie bazată pe un singur predictor, de exemplu estimarea reuşitei şcolare pe baza coeficientului de inteligență. Regresia multiliniară include mai mulți predictori în ecuație, precum nivelul de inteligență, motivația şcolară, nivelul de educație al părinților, pentru a putea estima mai bine reuşita şcolară. O altă denumire a regresiei multiliniare este **regresia liniară multiplă**. Cu siguranță că modelele multiliniare vor prezice întotdeauna mai bine decât modelele simple, care pornesc de la un singur predictor. Ideea este intuită şi în astrologie, unde pentru a prezice stilul de personalitate al unei persoane este necesar ca „specialistul" să cunoască mai mulți predictori: ziua de naştere, ora la care s-a născut, anul naşterii etc.

În continuare, pentru a putea înțelege mai bine modul de funcționare a regresiei, vom explica mai întâi regresia liniară simplă, bazată pe un singur

predictor, nu înainte de a preciza o informație capitală: regresia liniară poate fi aplicată doar dacă criteriul (variabila dependentă) a fost măsurat printr-o scală numerică (de tip interval sau proporții) şi dacă relația dintre predictori şi criteriu este liniară. Pentru alt gen de scale există alte tipuri de regresie precum, cea logistică (variabila dependentă dihotomică, de gen „Da" sau „Nu"), cea logistică multinominală (variabila dependentă este măsurată prin scală nominală, alta decât cea de tip dihotomic), sau cea ordinală (variabila dependentă este măsurată prin scală ordinală, de genul răspunsurilor în patru trepte: „nesatisfăcător", „satisfăcător", „bine" sau „foarte bine").

6.1. Regresia liniară simplă

6.1.1. Principii ale regresiei liniare

Imaginați-vă că sunteți un proaspăt absolvent de psihologie ajuns să lucreze în departamentul de resurse umane din cadrul unei mari companii. În această calitate, sunteți pus să selectați dintre mai mulți candidați pe cei care vi se par mai potriviți pentru a lucra în departamentul vânzări, în prima linie a contactului direct cu potențialii clienți. Care dintre candidați este mai potrivit pentru a fi angajat? Pentru a răspunde la întrebare vom prezenta trei alternative posibile.

Situația I. Ați făcut un curs introductiv de statistică, dar nu v-a plăcut.

În acest caz, cel mai probabil, veți aborda intuitiv situația „ghicind" performanța candidaților pe baza unui set de întrebări sau indicatori gândiți ad-hoc, fără prea mare succes.

Situația II. Ați făcut un curs serios, dar introductiv de statistică, fără a include regresia.

Într-o asemenea situație, veți putea estima ceva mai corect performanțele viitoare ale candidaților pornind de la proprietățile distribuției normale a datelor. Dacă performanța era exprimată în volumul vânzărilor dintr-o lună, am putea estima că aproximativ 68% din candidați vor obține performanțe în jurul mediei (într-un interval situat la o abatere standard distanță de o parte şi de alta a mediei) (vezi secțiunea 2.6.2. *Proprietățile unei distribuții normale* din cadrul capitolului I).

Situația III. Ați făcut un curs serios de statistică şi intuiți că regresia poate fi soluția corectă.

Într-adevăr, prin regresie se pot estima rezultatele viitoare ale unor persoane pe baza unor indicatori din prezent. Aşa cum pe baza coeficientului de inteligență se poate prevedea într-o anumită măsură evoluția şcolară a elevilor şi în cazul vânzărilor, se poate prevedea performanța pe baza unor indicatori care au legătură cu acest aspect. Totul porneşte de la existența unei asocieri între indicatorul utilizat pentru a estima performanța viitoare şi criteriul ales pentru a măsura această performanță. De exemplu, dacă există o corelație pozitivă între

extravertire şi performanța din vânzări (figura 6.1.), se poate estima că persoanele cu un nivel ridicat de extroversiune tind să obțină performanțe superioare în vânzări, fiind mai potrivite să ocupe asemenea poziții decât persoanele introvertite.

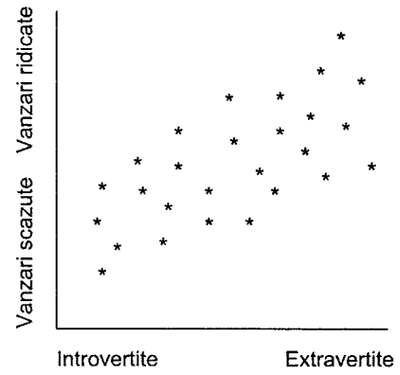


Figura 6.1. *Norul de puncte vizând relația dintre nivelul de extroversiune şi volumul vânzărilor*

Pornind de la această asociere se poate estima că persoanele cu un nivel ridicat de extravertire vor fi avantajate în domeniul vânzărilor, deoarece tind să obțină un număr mai mare de vânzări.

Regresia urmăreşte să găsească o ecuație care să ne ajute să estimăm cât mai exact valoarea performanței viitoare. Dacă corelația dintre două variabile ar fi egală cu 1 sau cu -1 am putea anticipa exact rezultatele viitoare, deoarece linia de regresie obținută reuşeşte să treacă prin toate rezultatele. Aceasta ar însemna că nu există nici o diferență între rezultatele constatate şi estimările făcute.

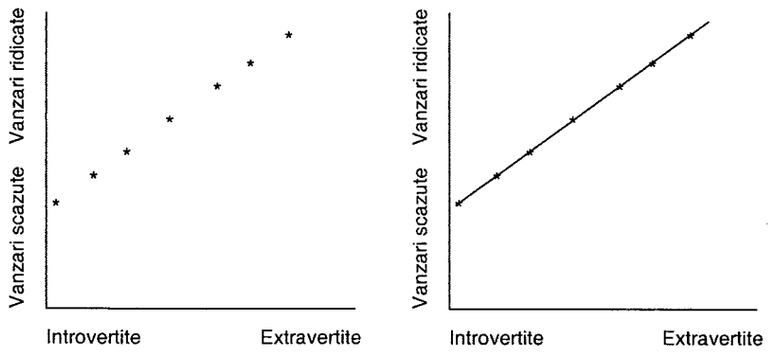


Figura 6.2. *Norul de puncte într-o situație de corelație ideală (stânga) unit prin dreapta de regresie (dreapta)*

Din nefericire, situația prezentată în figura 6.2 nu este reală, deoarece, în practică, nu se obțin coeficienți de corelație de +1 sau -1. De aceea, prin regresie se urmăreşte găsirea acelei linii care să conducă la o estimare cât mai corectă a rezultatelor. Cu alte cuvinte, se doreşte găsirea acelei linii de regresie care să reproducă cel mai bine direcția de evoluție a norului de puncte.

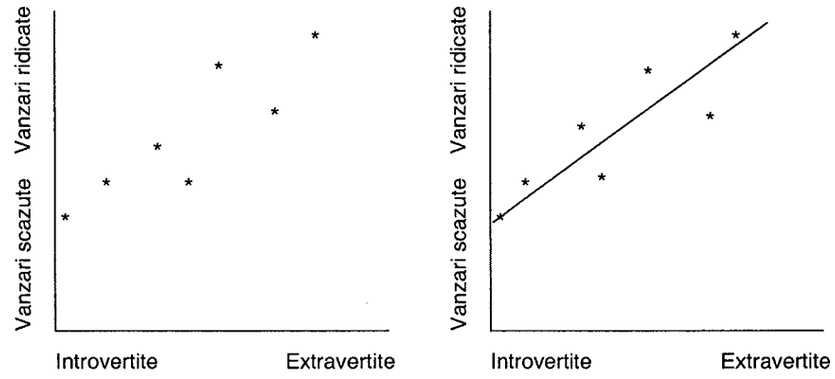


Figura 6.3. *Norul de puncte într-o situație de corelație probabilă (stânga) unit prin dreapta de regresie (dreapta)*

Pentru a estima prin regresie rezultatele viitoare ale unui anumit criteriu este necesar ca predictorul respectiv să coreleze destul de bine cu criteriul care trebuie estimat. Dacă nu există corelație între cele două elemente, linia de regresie nu poate fi estimată acceptabil, deoarece poate lua direcții diferite.

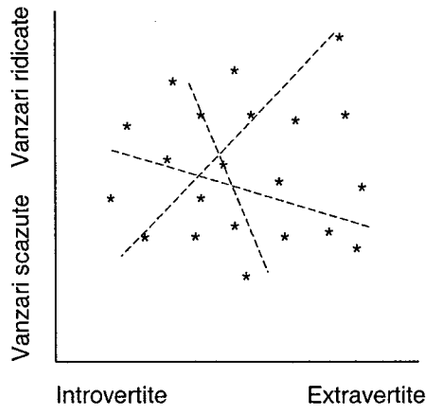


Figura 6.4. *Direcții multiple posibile ale liniei de regresie în absența corelației dintre predictor şi criteriu*

Pentru a stabili linia de regresie se apelează la proprietățile matematice ale unei drepte, deoarece orice dreaptă poate fi determinată prin următoarea ecuație:

$$y = b_0 + b_1X \qquad (6.1)$$

unde b0 este punctul de intersecție al ordonatei (interceptul), iar b1 este panta de regresie (cu cât creşte y, atunci când x se modifică cu o unitate).

Panta de regresie ia o valoare pozitivă, dacă există o corelație pozitivă între x şi y şi o valoare negativă, dacă există o corelație negativă între cele două variabile. Pentru a uşura înțelegerea acestui demers putem inspecta vizual cele două imagini din figura 6.5. Imaginea din stânga are drept ecuație de regresie: y = 2,5 + 0,5*X, în timp ce imaginea din dreapta are ca ecuație de regresie: y = 8 – 1,5*X

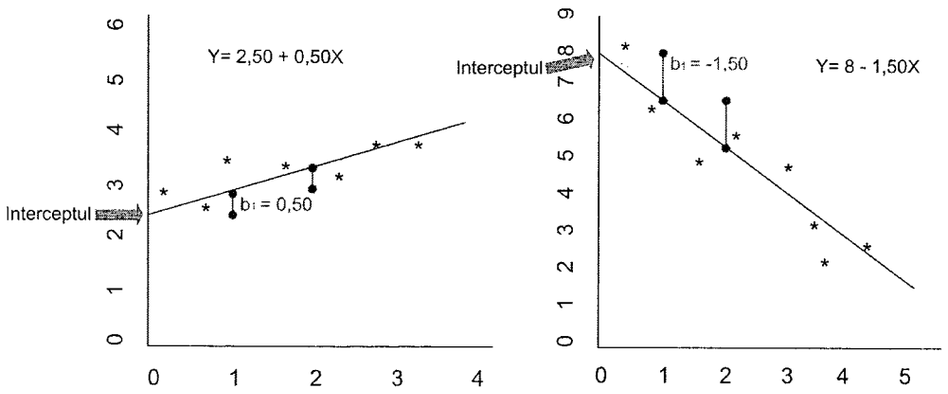


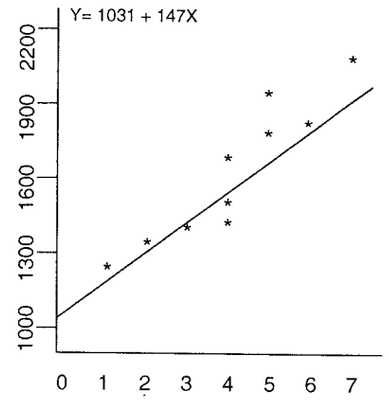
Figura 6.5. *Două exemple ipotetice de linie de regresie*

Pe baza imaginilor anterioare se poate observa mai bine cum se reprezintă grafic linia de regresie pe baza ecuației obținute. De exemplu, în imaginea din stânga, punctul de intersecție a liniei de regresie cu ordonata este situat la valoarea de b0 = 2,50, în timp ce în dreapta el ia valoarea 8 (b0 = 8). Ele reprezintă estimările lui Y, în situația în care X ia valoarea 0. Transpunând în cazul concret al vânzărilor, dacă pe abscisă este trecut numărul de itemi vizând extrovertirea, iar pe ordonată este trecut numărul de maşini vândute, din imaginea din stânga am putea estima că o persoană care a obținut scorul zero la extravertire va reuşi să vândă pe lună 2,5 maşini (y = 2,5 + 0,5*0), iar o persoană care a obținut scorul trei la extravertire va reuşi să vândă 4 maşini pe lună (y = 2,5 + 0,5*3).

De asemenea, se observă din cele două imagini că panta dreptei este mai abruptă în cazul imaginii din dreapta, deoarece valoarea b1 = -1,50 este mai mare

în valoare absolută decât b1 = 0,50 din imaginea din stânga. Valoarea acestei pante indică cu cât creşte (în cazul valorilor pozitive) sau cu cât scade (în cazul valorilor negative) Y atunci când X creşte cu o unitate.

Linia de regresie care redă cel mai bine norul de puncte nu se stabileşte pe baza unor criterii subiective, cum ar fi simpla inspectare vizuală. Metoda utilizată cel mai frecvent pentru a calcula interceptul şi panta de regresie se numeşte metoda celor mai mici pătrate (în engleză „least squares"). Metoda se bazează pe calcularea sumei pătratelor diferențelor dintre valorile observate şi valorile estimate. Ne vom baza pe datele din figura 6.6 pentru a explica metoda amintită.



X	Y	Y*	Y-Y*	$(Y-Y^*)^2$
4	1500	1619	-119	14161
5	1800	1766	34	1156
6	1850	1913	-63	3969
3	1400	1485	-85	7225
2	1350	1325	25	625
4	1700	1619	81	6561
7	2100	2060	40	1600
5	1970	1766	204	41616
4	1420	1619	-199	39601
1	1260	1178	82	6724

Suma = 0 Suma = 123.238

Legenda: X - Nivel de extravertire
Y - Vanzari lunare realizate (in Euro)
Y* - Vanzari lunare estimate (in Euro)

Figura 6.6. *Linia de regresie vizând relația dintre scorul la extravertire şi valoarea vânzărilor lunare în Euro*

Pe baza ecuației de regresie se estimează rezultate uşor diferite decât cele constatate. De exemplu, primul subiect a realizat vânzări lunare de 1500 de EURO, în condițiile în care se estimau, pe baza ecuației de regresie, vânzări lunare în valoare de 1619 EURO. Parametrii ecuației de regresie se stabilesc, astfel încât suma pătratului diferențelor dintre valorile Y observate şi cele estimate să ia valoarea cea mai mică (ultima coloană din tabel). De exemplu, dacă am fi schimbat ecuația de regresie în Y = 1050 + 150X, suma pătratelor diferențelor ar fi fost mai mare, mai exact 130.400. Prin urmare, în cazul de față valorile b0 de 1031 şi b1 de 147 sunt oferite de către programele statistice pentru că duc la cea mai mică sumă a pătratelor diferențelor dintre cele două seturi de date (valorile observate şi valorile estimate). Cu alte cuvinte, parametrii ecuației de regresie rezultați asigură soluția cea mai apropiată de rezultatele reale observate.

Pentru a decide dacă o ecuație de regresie descrie bine setul de date, spre a putea fi folosită în scop predictiv, trebuie să decidem cât de mare poate fi

diferența dintre rezultatele observate şi cele estimate. În acest sens, există o serie de parametri, valabili atât în cazul regresiei liniare simple, cât şi în cazul regresiei multiliniare, pentru a descrie eficiența ecuației de regresie în estimarea rezultatelor.

6.1.2. Evaluarea eficienței unei ecuații de regresie

Între indicatorii care ne oferă informații asupra eficienței unui model de regresie vom aminti testul F (întâlnit şi în cazul ANOVA) şi R^2 (coeficient de determinare întâlnit şi în cazul corelației).

Primul ne arată în ce măsură există diferențe semnificative statistic între estimările oferite pe baza ecuației de regresie implicate în comparație cu estimările bazate pe valorile mediei. Dacă predictorii incluşi în ecuația de regresie conduc la estimări semnificativ mai bune decât estimările bazate pe rezultatele medii, atunci valoarea F va fi semnificativă statistic. Dacă F nu este semnificativ statistic, atunci ecuația de regresie nu-şi are rostul, deoarece ea nu poate estima cu nimic mai bine decât simpla consultare a mediei răspunsurilor subiecților la criteriu.

Cel de-al doilea parametru poate fi consultat dacă F este semnificativ statistic. După Field (2000), R^2 oferă informații privitoare la procentajul din dispersia variabilei criteriu care poate fi explicată prin modul de evoluție a predictorilor. În cazul regresiei liniare simple, R^2 este coeficientul de determinare, acelaşi din cazul corelației simple, în timp ce în cazul regresiei liniare multiple, R^2 este coeficientul de determinare multiplă. Deşi are acelaşi principiu de interpretare, coeficientul de determinare multiplă reprezintă procentul din dispersia criteriului explicată de acțiunea comună a tuturor predictorilor implicați.

În cazul datelor prezentate în figura 6.6, SPSS-ul oferă următoarele rezultate:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,915[a]	,838	,818	123,074

a. Predictors: (Constant), Nivel de extravertire

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	626471,5	1	626471,453	41,359	,000[a]
	Residual	121178,5	8	15147,318		
	Total	747650,0	9			

a. Predictors: (Constant), Nivel de extravertire

b. Dependent Variable: Vanzari lunare in Euro

Figura 6.7. *Output oferit de SPSS 11.0*
SPSS este o marcă înregistrată.

Din cel de-al doilea tabel se observă că valoarea F este semnificativă statistic, fapt ce induce ideea superiorității ecuației de regresie, bazată pe introducerea nivelului de extravertire ca predictor, în estimarea vânzărilor în comparație cu estimarea vânzărilor pe baza mediei valorilor observate. Mai mult, consultând primul tabel, se observă că nivelul de extravertire al unei persoane explică 83,8% din variația vânzărilor observate (R^2 = 0,838). Cu alte cuvinte, dac• am încerca să explicăm diferențele constatate între persoane cu privire la vânzările lunare realizate de acestea, putem spune că o mare parte se datorează diferențelor în nivelul de extravertire dintre acestea. Restul de 16,2% din variația vânzărilor lunare înregistrate se datorează altor factori, neluați în seamă de modelul testat, care a inclus doar extroversiunea ca predictor. Aşadar, putem afirma că nivelul de vânzări realizate depinde şi de alți factori, deşi extroversiunea pare a fi variabila principală pe baza căreia putem estima nivelul de vânzări realizate. Din nefericire, asemenea procentaje ridicate, precum cel prezentat în problema anterioară nu sunt întâlnite în ştiințele sociale decât în exemple ipotetice. De obicei, R^2 are valori mult mai modeste, fiind cuprins destul de des între .10 şi .50, ceea ce înseamnă că majoritatea modelelor estimative din psihologie, bazate pe regresie explică (estimează) între 10% şi 50% din evoluția criteriului de interes.

După cum se observă din figura 6.8, valorile R^2 sunt mai mari, dacă linia de regresie descrie bine direcția evoluției norului de puncte (R^2 = .88 în imaginea din stânga) şi mai mici, dacă linia de regresie nu descrie adecvat norul de puncte, acestea fiind mai depărtate de direcția estimată (R^2 = .45 în imaginea din dreapta).

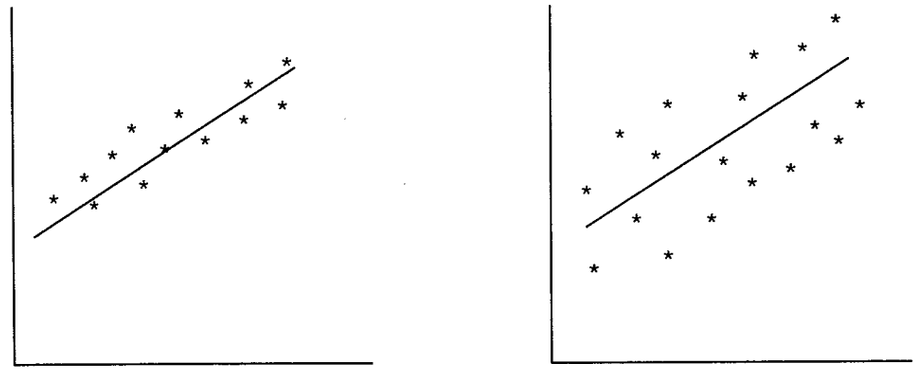


Figura 6.8. *Relația dintre gradul de împrăştiere a punctelor din nor şi valorile R^2*

Din nefericire, aşa cum remarcau Hankins, French şi Horne (2000), un model nu este adecvat doar pentru simplul fapt că R^2 este mare sau inadecvat pentru că R^2 este mic. Pentru a ne asigura că valoarea lui R^2 este apreciată corect ar trebui să facem o analiză a reziduurilor, adică a diferențelor dintre valorile observate şi cele estimate pe baza ecuației de regresie. Analiza este necesară

deoarece încălcarea unor condiții simple (ex. relație neliniară, heteroscedasticitate, prezența datelor neobişnuite) poate conduce la distorsiuni grave ale coeficientului de determinare. Despre aceste aspecte vom discuta în detaliu în secțiunea destinată condițiilor necesare de aplicare a regresiei liniare.

Revenind la figura 6.7, observăm şi alți indicatori relevanți. De pildă, în primul tabel, sub titulatura „R" este trecută valoarea coeficientului de corelație liniară între predictor şi criteriu, în cazul de față între nivelul de extravertire şi nivelul vânzărilor lunare realizate. În cazul regresiei liniare simple, această valoare indică, de fapt, coeficientul de corelație r Bravais-Pearson, între predictor şi criteriu. În cazul unei regresii multiliniare, R indică corelația multiplă dintre criteriu şi grupul de predictori incluşi. De asemenea, merită să fie remarcat faptul că R^2 reprezintă pătratul acestor coeficienți de corelație.

Un alt parametru regăsit în outputul SPSS este R^2 ajustat. Acesta încearcă să elimine unele dintre neajunsurile lui R^2. Acesta din urmă este dependent de numărul de predictori incluşi în ecuația de regresie şi de numărul de subiecți testați. În forma sa ajustată, R^2 încearcă să atenueze această influență, oferind un parametru mai puțin distorsionat. Totuşi, trebuie precizat că indicatorul R^2 ajustat oferit de SPSS nu are rol de contravalidare, aşa cum mulți specialişti ar fi tentați să creadă (Field, 2000). Astfel, el nu estimează procentajul din dispersia criteriului explicată de predictori în cazul în care datele ar fi extrase pe baza răspunsurilor subiecților dintr-un alt eşantion, ci arată doar eficiența modelului estimativ propus pentru eşantionul de subiecți testați, într-un mod cât mai puțin distorsionat de numărul de predictori sau de volumul eşantionului. Deşi oferă informații mai exacte decât coeficientul de determinare, R^2 ajustat rămâne sensibil la distorsiune în cazul nerespectării condițiilor necesare aplicării regresiei liniare.

6.1.3. Testarea individuală a eficienței predictorilor în estimarea evoluției criteriului

În cazul unei regresii liniare simple, dacă F a fost semnificativ statistic înseamnă că predictorul inclus, în situația de față nivelul de extravertire, poate fi utilizat în estimarea performanțelor din domeniul vânzărilor. În schimb, în cazul unei regresii multiliniare, eficiența per ansamblu a modelului nu înseamnă că fiecare dintre predictorii incluşi poate estima eficient evoluția criteriului. De pildă, în varianta unei regresii cu trei predictori, modelul poate fi considerat eficient, adică mai bun decât o estimare pe baza valorii medii a criteriului, chiar dacă numai un singur predictor are legătură cu criteriul. Pentru a vedea care dintre predictori poate estima cât mai corect evoluția criteriului, SPSS-ul oferă un alt output, pe baza căruia se poate exprima şi ecuația de regresie.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1031,249	101,614		10,150	,000
	Nivel de extravertire	147,232	22,894	,915	6,431	,000

a. Dependent Variable: Vanzari lunare in Euro

Figura 6.9. *Output oferit de SPSS 11.0 cu valorile coeficienților din ecuația de regresie*
SPSS este o marcă înregistrată.

Din această figură se poate deduce ecuația de regresie care poate fi scrisă sub forma:

Vânzări lunare în EURO = 1031,2 + 147,2*Nivel extravertire;

sau, renunțând la zecimale:

Vânzări lunare în EURO = 1031 + 147*Nivel de extravertire.

Pe baza acestei ecuații putem estima vânzările pe care le vor realiza alte persoane cu caracteristici similare persoanelor testate. De exemplu, o persoană care nu ar bifa nici un item ce exprimă extroversiunea va realiza vânzări de aproximativ 1031 de EURO lunar (a), în timp ce o persoană care a obținut scorul 7 la nivelul de extroversiune va realiza vânzări lunare duble, de aproximativ 2060 de EURO (b).

Aceste rezultate au fost obținute înlocuind în ecuația de regresie de mai sus, valorile observate ale extroversiunii pentru cele două persoane:

(a) Vânzări lunare în EURO = 1031 + 147*0
(b) Vânzări lunare în EURO = 1031 + 147*7

Cu alte cuvinte, valorile b reprezintă coeficienții nestandardizați ai ecuației de regresie, pe baza cărora estimăm cu cât creşte criteriul atunci când predictorul se măreşte cu o unitate. În exemplul dat, pornind de la faptul că b extravertire este 147, putem estima că volumul vânzărilor realizate de o persoană care a obținut, să zicem, cinci puncte la testul de extroversiune va avea vânzări mai mari cu 147 de EURO decât o persoană care a obținut patru puncte la testul de extroversiune, iar aceasta, la rândul ei, va obține cu 147 de EURO mai mult decât o persoană care are doar trei puncte la testul de extroversiune. În plus, interceptul denumit în SPSS „constant" exprimă nivelul vânzărilor obținute de o persoană în condițiile în care predictorul ia valoarea zero, în situația dată, dacă persoana obține zero puncte la testul de extroversiune.

Se poate observa că această evoluție se menține atunci când coeficientul de regresie al predictorului ales are valori pozitive. Dacă am fi obținut un b negativ, să presupunem b = -147, ar fi însemnat că volumul vânzărilor lunare scade cu 147 de EURO de fiecare dată când scorul la extroversiune creşte cu o unitate. Situația ar fi fost similară cu cea întâlnită în cazul unei corelații negative: când X creşte, Y tinde să scadă.

Totuşi, valorile nestandardizate b sunt dependente de scala de măsurare a predictorilor şi a criteriului. De exemplu, valoarea coeficientului b pentru testul de extroversiune se modifică, dacă extroversiunea este măsurată printr-o altă scală de măsurare sau, dacă vânzările lunare sunt exprimate în lei sau dolari şi nu în euro. Astfel, dacă vom transforma valorile exprimate în euro în lei, utilizând drept rată de schimb 40.000 lei = 1 EURO, vom obține alți parametri ai ecuației de regresie, deşi eficiența modelului rămâne aceeaşi.

Volum lunar de vânzări în lei = 41.253.980 + 5.889.273*Nivel extravertire

Ce importanță are acest lucru?

Prin faptul că parametrii nestandardizați sunt dependenți de scala de măsurare, rezultă imposibilitatea de a compara direct influența a doi sau mai multor predictori asupra criteriului. De exemplu, nu se poate compara direct dacă influența predictorului nivel de extroversiune asupra nivelului de vânzări (b1 = 147) este mai mare decât influența vârstei subiecților asupra volumului de vânzări (b2 = 18). Pentru a observa care predictor ajută mai mult la estimarea evoluției criteriului, se obişnuieşte compararea valorilor standardizate ale acestor coeficienți b, denumite, inclusiv de SPSS, beta.

Din figura 6.9 observăm că valoarea β pentru extravertire este .91, ceea ce înseamnă că, atunci când nivelul de extroversiune creşte cu o abatere standard, criteriul, adică volumul vânzărilor, creşte cu 0,91 abateri standard. Ecuația de regresie standardizată are interceptul egal cu zero, ceea ce face ca ea să poată fi scrisă sub forma:

Z vânzări lunare în EURO = .91*Z nivelul de extroversiune

După cum se observă, o asemenea interpretare nu este la fel de atrăgătoare în comparație cu cea bazată pe coeficienții nestandardizați, dar are avantajul esențial, în cazul regresiilor multiliniare, de a putea compara simultan eficiența predictorilor implicați în ecuația de regresie. Astfel, dacă am presupune că β vârstă este 1,21, am putea afirma că vârsta unei persoane este un predictor mai bun decât nivelul de extroversiune în estimarea vânzărilor lunare realizate. Acest lucru poate fi făcut, deoarece toate datele standardizate în cote z au aceeaşi unitate de măsură: media este zero, iar abaterea standard este egală cu unu.

O altă informație extrasă din figura 6.9 constă în faptul că ambii parametri, atât interceptul cât şi panta de regresie, sunt semnificativi statistic. În cazul

regresiei liniare simple, acest lucru era prevăzut prin faptul că valoarea F pentru întregul model era semnificativă statistic. În schimb, în cazul existenței mai multor predictori, existența unui F semnificativ statistic indică faptul că cel puțin un predictor poate estima evoluția criteriului, fără a porni de la ideea că toți predictorii vor avea capacitatea de a estima evoluția criteriului. În asemenea situații, este necesară consultarea valorilor t din dreptul fiecărui parametru prezentat în figura 6.9. pentru a vedea care dintre predictori explică semnificativ statistic evoluția criteriului. Interpretarea este simplă: dacă valoarea t asociată unui parametru este semnificativă statistic (ex. t = 6.41, p < .001 în cazul extroversiunii), atunci predictorul respectiv îşi aduce o contribuție semnificativă statistic la estimarea evoluției criteriului, în cazul respectiv a cifrei de vânzări lunare realizate.

Pentru a înțelege mai bine aceste detalii tehnice în interpretarea regresiei liniare, vom prezenta un exemplu de regresie multiliniară. Înainte de a face acest lucru, se cuvine să trecem în revistă, pe scurt, principalele modalități tehnice de obținere a ecuației de regresie în cazul celei de tip multiliniar.

6.2. Modalități de analiză în regresia multiliniară

De cele mai multe ori, modelele estimative bazate pe un singur predictor nu sunt la fel de eficiente ca cele care includ mai mulți predictori. De exemplu, un psiholog poate considera că o persoană suferă de depresie, dacă are mai mulți indicatori (ex. stimă de sine scăzută, scăderea apetitului alimentar şi sexual, preferința pentru culori închise etc.) decât, dacă ar lua în calcul un singur element. În situația în care sunt incluşi mai mulți indicatori în modelul estimativ, vorbim de o regresie multiplă. Analiza predictorilor incluşi într-o asemenea ecuație de regresie, indiferent de scopul acesteia, se poate face prin patru modalități de bază: metoda simultană, metoda ierarhică, metoda paşilor şi metoda celui mai bun set posibil (Licht, 1995).

Metoda simultană

Metoda simultană presupune introducerea concomitentă a tuturor predictorilor de interes în cadrul ecuației de regresie. De exemplu, dacă vrem să vedem în ce măsură sexul persoanei, vârsta acesteia, nivelul de educație şi unele trăsături de personalitate (ex. neuroticism şi extroversiune) pot prezice atitudinea subiecților față de telenovele, am putea realiza o singură ecuație de regresie, de forma:

Atitudine telenovele = b0 + b1*Sex + b2*Vârstă + b3*Educație + b4*Neuroticism + b5*Extroversiune

Această metodă este cel mai frecvent întâlnită, fiind opțiunea standard în SPSS.

Metoda ierarhică

Modalitatea presupune realizarea a cel puțin două analize de regresie, aflate în relație de tip cuib (toți predictorii utilizați într-unul dintre modele se găsesc în cel de-al doilea). Pornind de la exemplul anterior, dacă suntem interesați să vedem în ce măsură trăsăturile de personalitate extroversiune şi neuroticism aduc un plus predictiv în estimarea atitudinii față de telenovele, după eliminarea influenței factorilor demografici, vom avea următoarele două ecuații de regresie:

Atitudine telenovele = b0 + b1*Sex + b2*Vârstă + b3*Educație

Atitudine telenovele = b0 + b1*Sex + b2*Vârstă + b3*Educație + b4*Neuroticism + b5*Extroversiune

Fiecărei ecuații îi corespunde o anumită valoare F şi un anumit coeficient de determinare ajustat sau nu (R^2 ajustat sau R^2). Dacă nu există diferențe între cele două ecuații, în privința acestor valori, înseamnă că neuroticismul, respectiv extroversiunea, nu aduc un plus de informație cu privire la estimarea atitudinii subiecților față de telenovele. În schimb, dacă F schimbare, obținut prin diferența dintre cele două modele este semnificativ statistic, înseamnă că trăsăturile de personalitate aduc un plus de informații în estimarea atitudinii față de telenovele. În interpretare, acest lucru ar putea fi exprimat astfel: în condițiile în care persoanele au aceeaşi vârstă, acelaşi nivel de educație şi acelaşi sex, diferențele de personalitate la nivel de extroversiune şi neuroticism contribuie semnificativ statistic la tipul de atitudine adoptată față de telenovele. Din acest motiv, regresia ierarhică este frecvent utilizată în cazul regresiilor cu scop explicativ însă ea apare foarte rar în regresiile cu scop predictiv.

Metoda pașilor

Cunoscută în limba engleză sub titulatura de „stepwise", fie în variantă „forward", fie în variantă „backward", metoda este criticată intens de specialiştii în statistică, deşi se bucură de o anumită popularitate în tagma cercetătorilor fără o experiență statistică deosebită.

În varianta „înainte" („forward"), ea constă în selectarea din setul de predictori a aceluia care contribuie cel mai mult la estimarea criteriului, pe bază pur statistică. În exemplul dat, prin această metodă se realizează, ca prim pas, o analiză de regresie simplă, introducând separat fiecare dintre cei cinci predictori (sexul persoanei, vârsta, educația, extroversiunea şi neuroticismul), pentru a vedea influența lor în estimarea atitudinii față de telenovele. În urma acestui pas este selectat acel predictor care contribuie cel mai mult, prin valoarea coeficientului de determinare, la estimarea criteriului. Să presupunem că dintre cei cinci predictori, cea mai mare influență o exercită nivelul de educație. Această variabilă este selectată şi introdusă în următorul pas. În acest de-al doilea pas se compară modelele de regresie care au inclusă această variabilă şi o alta, dintre cele rămase pentru a fi testate. În cazul particular exprimat, în acest pas se compară patru modele posibile rămase:

Atitudinea față de telenovele = b0 + b1*Educație + b2*Sex (1)

Atitudinea față de telenovele = b0 + b1*Educație + b2*Vârsta (2)

Atitudinea față de telenovele = b0 + b1*Educație + b2*Extroversiune (3)

Atitudinea față de telenovele = b0 + b1*Educație + b2*Neuroticism (4)

Dintre acestea, se alege modelul care are cel mai mare coeficient de determinare multiplă. Să presupunem că dintre cele patru modele, primul are coeficientul de determinare multiplă cel mai mare. Dacă acest model este semnificativ mai bun decât modelul predictiv din pasul unu, bazat doar pe includerea nivelului de educație, el devine modelul cel mai bun şi se trece la pasul al treilea, în care se compară următoarele modele rămase:

Atitudinea față de telenovele = b0 + b1*Educație + b2*Sex + b3*Vârsta (1)

Atitudinea față de telenovele = b0 + b1*Educație + b2*Sex + b3*Extroversiune (2)

Atitudinea față de telenovele = b0 + b1*Educație + b2*Sex + b3*Neuroticim (3)

Să presupunem că cel mai bun model dintre acestea este cel de-al doilea, în care este inclusă extroversiunea. Acest model se compară cu modelul de la pasul 2, iar dacă este semnificativ mai bun, el devine cel mai bun model şi se trece la pasul patru, în care vor rămâne doar două modele de comparat, după algoritmul prezentat în primii paşi. Demersul poate continua până la epuizarea tuturor variabilelor, caz în care cel mai bun model este acela în care sunt incluşi toți cei cinci predictori. Dacă, însă, diferența dintre cel mai bun model din pasul trei şi cel mai bun model din pasul doi nu este semnificativă statistic, vom concluziona că nici unul dintre cele trei modele de mai sus nu poate fi considerat mai bun decât cel obținut la pasul doi, iar modelul final poate fi cel obținut la pasul doi:

Atitudinea față de telenovele = b0 + b1*Educație + b2*Sex

Prin urmare, prin metoda paşilor se încearcă găsirea pe bază exclusiv empirică a modelului care estimează cel mai bine evoluția datelor. În varianta „înapoi" („backward"), metoda paşilor porneşte invers, în pasul unu fiind prezentat modelul cu toate cele cinci variabile, în pasul doi fiind prezentate toate combinațiile posibile de modele cu patru predictori. Dacă vreunul dintre acestea nu este semnificativ diferit de modelul cu cinci predictori, el devine cel mai bun model posibil şi se trece la pasul trei, de comparare a acestui model, cu cel mai bun model obținut din combinarea a trei dintre cei patru predictori rămaşi etc.

Deşi, în aparență metoda paşilor este atractivă, prin extragerea acelui număr de predictori care explică cel mai bine evoluția criteriului, ea suferă diverse limite statistice şi metodologice. Din punct de vedere metodologic, apare pericolul extragerii unor predictori irelevanți sau al renunțării la unii predictori relevanți sub aspect teoretic. Din punct de vedere statistic, cresc şansele de a obține un model specific, aplicabil doar pe eşantionul testat, ce nu poate fi replicat pe alți subiecți.

La aceste neajunsuri se adaugă şi unele greşeli de atribuire a gradelor de libertate în cazul regresiei prin metoda paşilor, fapt ce creşte probabilitatea de a comite eroarea de tip I, care, în acest caz, ar consta în includerea unor predictori care nu ar fi trebuit incluşi (Pedhazur şi Schmelkin, 1991; Thompson, 1995a).

Metoda celui mai bun set posibil

Acest gen de regresie se apropie de principiul regresiei în paşi, fiind într-un fel, o generalizare a acesteia. Ea constă în compararea tuturor modelelor posibile în funcţie de numărul de predictori incluşi în model. De exemplu, dacă am include trei predictori şi am utiliza metoda celui mai bun set posibil, calculatorul va selecta modelul cu coeficientul de determinare ajustat cel mai mare dintre toate cele şapte modele posibile (modelul care păstrează toţi cei trei predictori, trei modele cu câte doi predictori, trei modele cu un singur predictor). Deoarece se bazează pe aceleaşi principii ca cele întâlnite în cazul regresiei în paşi, metoda are aceleaşi neajunsuri.

Prin urmare, deşi există mai multe metode de estimare a modelului predictiv cu cel mai ridicat nivel de eficienţă în estimarea criteriului, recomand evitarea metodei paşilor şi cea a celui mai bun set posibil. În funcţie de scopul urmărit, veţi avea de ales între metoda simultană şi metoda ierarhică, ambele posibil de selectat în SPSS. De altfel, metoda simultană este şi opţiunea standard din SPSS (în engleză „enter"). Pentru a utiliza regresia ierarhică, vom menţine metoda „enter" însă după introducerea primului set de variabile vom apăsa butonul „next", iar apoi vom introduce atât variabilele din primul set, cât şi cele care dorim să le introducem în cel de-al doilea.

6.3. Regresia multilinară în scop predictiv

La fel cum „specialiştii" în astrologie consideră că pot deduce mai bine personalitatea pe baza mai multor predictori (ex. ziua de naştere, ora naşterii, locul naşterii) decât pe baza unui singur criteriu, şi în cazul predicţiei pe bază de regresie sunt preferate modelele multiple (cu mai mulţi predictori) în faţa celor simple (cu un singur predictor). Dacă revenim la exemplul prezentat în secţiunea destinată regresiei simple, ce viza găsirea unor predictori care să ne ajute să estimăm cât mai bine performanţa în vânzări a candidaţilor testaţi, ar fi bine să ţinem seama şi de alte variabile, în afara extroversiunii. De exemplu, am putea presupune că performanţa în vânzări este afectată de experienţa profesională în domeniu („Câţi ani de experienţă profesională în domeniul vânzărilor au candidaţii?), gradul de atractivitate fizică a personalului din vânzări şi de gradul de neuroticism al acestora, cunoscându-se faptul că un nivel ridicat de neuroticism corespunde unei afectivităţi negative şi unei rezistenţe scăzute la frustrare şi stres.

Pentru a da mai multă concretețe acestei situații, vom presupune că 60 de persoane angajate de două luni în domeniul vânzărilor constituie lotul pe baza căruia se va estima ecuația de regresie vizând performanța în vânzări. Pe baza acestei ecuații vor putea fi apoi selectați viitorii angajați în domeniu. Datele sunt prezentate în fișierul regresie multiliniara.sav.

De asemenea, se cuvine să precizăm modul în care au fost operaționalizate variabilele utilizate pe post de predictori, precum şi criteriul. Acesta din urmă a fost reprezentat de venitul lunar obținut din vânzări, de fiecare angajat, în cea de-a doua lună de muncă. Dintre predictori, extroversiunea şi neuroticismul au fost obținute prin intermediul Inventarului de Personalitate Eysenck (EPI), scorurile obținute fiind cuprinse, pentru ambele scale, între 0 şi 24 de puncte. Gradul de atractivitate a fost obținut pe baza sincronizării răspunsurilor de la doi colegi de sex opus, fiind utilizată o scală de la 1 la 10, unde 1 înseamnă un aspect fizic neîngrijit şi neplăcut, iar 10 un aspect fizic îngrijit şi foarte plăcut. În privința experienței profesionale, aceasta a fost codată în trei categorii distincte. Persoanele care nu aveau nici un fel de experiență profesională anterioară au fost notate cu 0, cele care beneficiau de experiența profesională, dar nu în domeniul vânzărilor, au fost codate cu 1, iar cele care aveau experiență profesională în domeniul vânzărilor erau cotate cu 2. Se observă că, spre deosebire de ceilalți predictori şi de criteriu, care erau măsurați prin scale de tip numeric, experiența profesională a fost operaționalizată prin intermediul unei scale de tip nominal (în acest caz particular poate fi percepută şi ca o scală de tip ordinal). Pentru a analiza influența predictorilor de acest tip se recurge la un artificiu metodologic ce necesită transformarea variabilelor categoriale în variabile de tip „dummy".

Variabilele de tip „dummy"

Sub această denumire sunt specificate în regresie toate variabilele categoriale dihotomice, care permit răspunsuri de tip 1 sau 0 (Schroeder, Sjoquist şi Stephan, 1986). De exemplu categoriile „bărbat" şi „femeie" din cadrul variabilei „sexul persoanei"; categoriile „da" sau „nu" vizând o anumită atitudine sunt din start variabile de tip dummy. Una dintre categoriile de răspuns va fi notată cu 1 (ex. femeie = 1; da = 1), iar cealaltă categorie de răspuns va fi notată cu 0 (ex. bărbat = 0; nu = 0). Răspunsurile notate cu zero mai sunt denumite şi categorii de referință. Prin intermediul regresiei, coeficienții b rezultați arată cu cât se modifică criteriul atunci când predictorul creşte cu o unitate. În cazul particular al variabilelor dummy, b arată cu cât se modifică criteriul atunci când predictorul îşi schimbă categoria de răspuns (de la 0 la 1). De exemplu, dacă b pentru sexul persoanei ar fi 1000 şi am păstra drept criteriu volumul vânzărilor realizate în EURO, am putea spune că femeile (femeie = 1), în comparație cu bărbații (bărbat = 0), reuşesc să obțină cu 1000 de EURO vânzări mai mari în domeniu.

Dacă lucrurile sunt simple în cazul variabilelor categoriale de tip dihotomic, ele se complică puțin în cazul unor variabile categoriale cu mai mult de două trepte. Spre exemplu, dacă ar fi să ținem cont de starea civilă a unei persoane, ca

variabilă predictor, aceasta poate fi exprimată prin patru categorii de răspuns: 0 = necăsătorit (burlac); 1 = căsătorit; 2 = divorțat; 3 = văduv. Într-o asemenea situație, pentru a vedea efectul stării civile asupra unei variabile criteriu este necesară transformarea variabilei în variabile de tip dummy (dihotomice). Se vor forma un număr de k-1 variabile dummy, unde k este numărul de categorii de răspuns. În acest caz, dată fiind existența a patru categorii de răspuns vom obține trei variabile dummy. Ce vor avea în comun aceste variabile nou formate va fi categoria de referință. Alegerea acesteia se face în funcție de importanța relativă a răspunsurilor posibile şi de frecvența acestora. În cazul stării civile, am putea alege ca valoare de referință prima categorie (0 = necăsătorit), ceea ce înseamnă că cele trei variabile dummy rezultate vor fi: căsătorit, divorțat şi văduv. Pentru a obține noile variabile se recurge la următoarele transformări. Pentru variabila dummy „căsătorit" vom avea doar două răspunsuri posibile: 1, dacă persoana este căsătorită şi 0, dacă persoana este necăsătorită, divorțată sau văduvă. Pentru variabila dummy „divorțat" vom cota 1, dacă persoana este divorțată şi 0, dacă persoana este necăsătorită, căsătorită sau văduvă. În sfârşit, variabila dummy „văduv" are cota 1, dacă persoana este văduvă şi 0, dacă ea se află într-una dintre celelalte trei categorii.

Dezavantajul transformării variabilelor categoriale în variabile de tip dummy constă în creşterea artificială a numărului de predictori, însă această soluție este singura care permite introducerea variabilelor categoriale ca predictori în analiza de regresie.

Tabelul 6.1. Transformarea variabilei „experien" în variabile de tip dummy:

Număr Subiect	„Experien"	Dummy 1 „Exp_vanz"	Dummy 2 „Exp_altd"
1	0	0	0
2	1	0	1
3	2	1	0
4	1	0	1
5	0	0	0
6	2	1	0
7	1	0	1
8	2	1	0
9	1	0	1
10	0	0	0

Legendă categorie răspunsuri „experien":
0 – fără experiență profesională;
1 – experiență profesională în alt domeniu;
2 – experiență profesională în vânzări.

Dacă ați înțeles procesul de transformare a variabilelor categoriale în variabile de tip dummy, atunci veți intui corect transformările necesare în cazul exemplului dat, vizând variabila experiență profesională. Dintre cele trei categorii de răspunsuri: 2 - cu experiență profesională în domeniul vânzărilor; 1 – cu

experiență profesională în alte domenii; 0 – fără experiență profesională anterioară, am decis să alegem drept categorie de referință ultimul răspuns. Această decizie conduce la formarea a două variabile dummy, pe care le vom denumi exp_vanz şi exp_altd. Prima variabilă dummy va cota indivizii care au experiență în domeniul vânzărilor cu 1, iar pe toți ceilalți cu 0. Cea de-a doua variabilă dummy va cota indivizii care au experiență prefesională în alte domenii cu 1, iar pe toți ceilalți cu 0. Tabelul 6.1 exemplifică această transformare în cazul primilor zece subiecți din fişierul regresie multipla.sav.

În cazul regresiei liniare SPSS-ul nu oferă posibilitatea de a transforma o variabilă categorială în variabile dummy din meniu. Acest lucru poate fi făcut doar prin apelul la sintaxă. De aceea, cei care nu sunt familiarizați cu apelul la sintaxă vor trebui să recodeze variabila categorială în variabile dummy, fie manual, fie prin apelul la opțiunea disponibilă prin calea „Transform"-„Recode"-„Into different variables". În alte situații, precum în cazul regresiei logistice, SPSS-ul permite recodarea automată a variabilelor categoriale în variabile dummy. În acest caz, răspunsul ales drept categorie de referință trebuie să fie întotdeauna codat cu 0, pentru că SPSS alege automat această valoare ca referință. Dacă nici unul dintre răspunsuri nu este codat 0, SPSS-ul va alege ca valoare de referință categoria de răspuns cu valoarea cea mai mare. De exemplu, dacă răspunsurile la variabila „starea civilă" ar fi fost codate: 1 – necăsătorit; 2 – căsătorit; 3 – divorțat; 4 – văduv, SPSS-ul ar fi ales automat categoria „văduv" ca valoare de referință. În schimb, dacă răspunsurile sunt codate precum în exemplul dat: 0 – necăsătorit; 1 – căsătorit; 2 – divorțat; 3 – văduv, va alege drept categorie de referință varianta „necăsătorit".

Revenind la exemplul dat, vom avea de testat eficiența unui model predictiv alcătuit din cinci predictori asupra performanțelor în vânzări. Cei cinci predictori sunt extroversiunea, neuroticismul, gradul de atractivitate fizică, dummy „experiență în vânzări" şi dummy „experiență în alte domenii". Baza de date aferentă este denumită regresie multipla.sav.

Prin intermediul regresiei liniare, disponibile în SPSS prin calea „Analyze"-„Regression"-„Linear", vom obține răspunsul la trei întrebări importante:

1. este acest model eficient?
2. care dintre predictori contribuie la eficiența modelului?
3. care dintre predictori are cea mai mare influență asupra eficienței modelului?

Răspunsurile la cele trei întrebări pentru cazul dat se găsesc sintetizate în tabelele output oferite de SPSS 11.0. şi prezentate în figura 6.10.

Primele două tabele arată că cei cinci predictori incluşi conduc la un model de regresie semnificativ statistic mai bun decât cel obținut pe baza mediei – $F(5, 54) = 10.46$, $p < .001$, care este capabil să explice, în formă ajustată, o proporție de 44,5% (R^2 ajustat = .445) din evoluția dispersiei performanțelor în vânzări.

Restul de 55,5%, rămas neexplicat, se datorează altor factori, neintroduşi în modelul de regresie. Totuşi, merită menționat că o asemenea valoare, indică o

bună putere explicativă în cadrul ştiinţelor sociale, deoarece majoritatea modelelor de regresie din această arie se rezumă la coeficienţi de determinare multiplă cuprinşi între .10 şi .50.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,701[a]	,492	,445	209,850

a. Predictors: (Constant), Experienta in alte domenii, Grad de atractivitate fizica, Neuroticism, Extroversiune, Experienta in domeniul vanzarilor

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2302920	5	460583,993	10,459	,000[a]
	Residual	2377998	54	44036,996		
	Total	4680918	59			

a. Predictors: (Constant), Experienta in alte domenii, Grad de atractivitate fizica, Neuroticism, Extroversiune, Experienta in domeniul vanzarilor

b. Dependent Variable: Vanzari lunare in EURO

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1038,252	108,435		9,575	,000
	Extroversiune	24,472	5,830	,420	4,198	,000
	Neuroticism	-8,898	5,283	-,167	-1,684	,098
	Grad de atractivitate fizica	33,868	12,707	,266	2,665	,010
	Experienta in domeniul vanzarilor	217,933	76,369	,368	2,854	,006
	Experienta in alte domenii	13,661	72,104	,024	,189	,850

a. Dependent Variable: Vanzari lunare in EURO

Figura 6.10. *Output SPSS 11.0 pentru regresia multiplă*
SPSS este o marcă înregistrată.

6.3.1. Interpretarea coeficienţilor de regresie

Pentru a interpreta coeficienţii de regresie corespunzători fiecărui predictor introdus în ecuaţie, se ţine seama de trei criterii:

- este acesta semnificativ statistic?
- are o valoare pozitivă sau negativă?
- este predictorul important?

Semnificaţia statistică a predictorilor

Din ultimul tabel al figurii 6.10. se pot extrage informaţii preţioase legate de ultimele două întrebări formulate, şi anume: care dintre predictori contribuie la eficienţa modelului şi care este ponderea predictorilor în cadrul modelului.

La o primă vedere, pe baza testului de semnificaţie t, se constată că, excluzând interceptul, doar trei predictori din cinci contribuie semnificativ statistic la estimarea performanţelor din vânzări. Cei trei predictori sunt extroversiunea, nivelul de atractivitate fizică şi variabila dummy ce opune persoanele cu experienţă profesională în domeniul vânzărilor celor care nu aveau nici un fel de experienţă.

Totuşi, ţinând cont că alegerea predictorilor nu s-a făcut la întâmplare, ci pe baza unor premise teoretice, am putea spune că avem de a face cu o ipoteză de tip unilateral, în care se specifică din start direcţia de evoluţie a criteriului în funcţie de predictori. Astfel, ne-am aştepta ca persoanele care au un nivel de extroversiune şi de atractivitate fizică mai ridicat, care au experienţă profesională în domeniul vânzărilor şi care au un nivel scăzut de neuroticism, vor obţine performanţe superioare în domeniul vânzărilor faţă de persoanele care au caracteristici opuse. În asemenea condiţii, trebuie să ţinem seama că semnificaţia testului t oferită de SPSS implică o ipoteză de tip bilateral. În consecinţă, pragul de semnificaţie poate fi înjumătăţit pentru a testa semnificaţia ipotezei unilaterale. În urma consultării ultimului tabel prezentat în figura 10, se observă că predictorul neuroticism devine semnificativ statistic în estimarea evoluţiei performanţelor în vânzări, deoarece în urma înjumătăţirii pragului de semnificaţie bilateral .096 se obţine un $t = -1.68$, $p < .05$.

Semnul predictorilor

Dacă coeficientul de regresie b are o valoare pozitivă înseamnă că există o relaţie directă între predictor şi criteriu. Astfel, atunci când predictorul creşte cu o unitate, criteriul creşte de asemenea. În cazul unor valori negative ale coeficientul de regresie există o relaţie inversă între predictor şi criteriu: când predictorul creşte cu o unitate, valoarea criteriului scade. Semnul predictorilor se interpretează doar în cazul în care aceştia contribuie semnificativ statistic la evoluţia criteriului. În exemplul dat, nu ne interesează semnul ultimei variabile dummy, care face diferenţa dintre persoanele cu experienţă profesională în alte domenii şi cele fără experienţă profesională, deoarece în lipsa semnificaţiei statistice a acestui predictor nu se poate afirma că acesta ajută în vreun fel la estimarea performanţei în vânzări. În schimb, cu excepţia neuroticismului care are un coeficient de corelaţie nestandardizat negativ (b = -8,898), toţi ceilalţi au o relaţie directă cu criteriul.

Ponderea predictorilor

În forma lor nestandardizată, coeficienţii de regresie corespunzători fiecărui predictor arată cu cât se modifică criteriul, atunci când predictorii cresc cu o unitate. În cazul regresiei multiple, precum în problema de faţă, această

interpretare se păstrează, adăugându-se doar sintagma „în condițiile păstrării constante a celorlalți predictori".

De exemplu, în cazul extroversiunii, b a fost 24,47, ceea ce înseamnă că o persoană care a obținut, spre exemplu, scorul 11 la proba de extroversiune, va obține vânzări lunare mai mari cu 24,47 EURO decât o persoană care a obținut doar 10 puncte la proba de extroversiune. Această estimare este valabilă în condițiile în care păstrăm constantă influența celorlalte variabile, o asemenea situație ipotetică fiind cazul în care cele două persoane obțin rezultate identice la toate celelalte probe, singura diferență fiind faptul că la proba de extroversiune, una dintre ele obține un scor cu un punct mai mult decât cealaltă.

Similar, în cazul neuroticismului, interpretarea asociată rezultatului b = -8,898 ar fi: o persoană care obține la scala de neuroticism cu un punct mai mult decât o altă persoană, în condițiile în care la celelalte probe au obținut performanțe egale (ceilalți predictori au fost ținuți constanți), se estimează că va avea cu 8,89 EURO vânzări lunare mai mici decât cea de-a doua.

În cazul variabilelor dummy, interpretarea este uşor diferită. Astfel, în cazul variabilei dummy „exp_vanz", care are un coeficient nestandardizat de regresie semnificativ statistic (b = 217,93) se estimează diferențele în vânzări dintre o persoană care are rezultatul 0 la variabila dummy respectivă şi o altă persoană care are cota 1 la aceeaşi variabilă, în condițiile în care influența celorlalți predictori este ținută constantă. În exemplul dat, dacă două persoane obțin rezultate identice la nivelul de extroversiune, la cel de neuroticism şi la cel de atractivitate fizică, dar diferă în ceea ce priveşte experiența profesională, în sensul că una are experiență în domeniul vânzărilor (exp_vanz = 1), iar cealaltă nu are nici un fel de experiență (exp_vanz = 0), prima se estimează că va obține lunar, în vânzări, cu 217,93 EURO mai mult decât cea de-a doua. Cu alte cuvinte, în condițiile menținerii constante a influenței celorlalți predictori, experiența anterioară în domeniul vânzărilor aduce un plus lunar de 217,93 EURO în comparație cu absența oricărei experiențe profesionale.

Se observă în cadrul acestei variabile dummy că se face comparația cu categoria de referință aleasă, în cazul dat, persoanele fără nici un fel de experiență profesională. În cadrul acestei comparații nu se ține seama de o altă categorie de răspunsuri posibile, aceea a persoanelor care au experiență profesională, dar în alte domenii decât cel al vânzărilor. Comparația dintre aceste persoane şi cele fără experiență se obține prin consultarea celei de-a doua variabile dummy „exp_altd", al cărei coeficient de regresie a fost nesemnificativ statistic (b = 13,66). Se observă că prin alegerea categoriei de referință „persoane fără experiență profesională" se face de fapt compararea dintre aceasta şi toate celelalte categorii posibile. În schimb, este imposibilă compararea dintre categoriile care nu joacă rol de referință, în cazul de față fiind imposibil tehnic să comparăm diferențele în performanța din vânzări dintre persoanele care au experiență în acest domeniu şi cele care au experiență în alte domenii. Pentru a putea face acest lucru este necesar să alegem altă categorie de răspuns ca variabilă de referință. De exemplu, dacă am nota cu 0

persoanele care au experiență în domeniul vânzărilor, am putea compara această categorie de referință cu persoanele care au experiență în alte domenii şi cu cele care nu au nici un fel de experiență profesională. Similar, într-un astfel de scenariu, nu am putea compara persoanele care au experiență profesională în alte domenii cu cele care nu beneficiază de nici un fel de experiență.

Dacă se vor alege alte categorii de referință, trebuie să rețineți că valorile coeficienților standardizați se vor schimba, în asemenea situații fiind necesară o nouă analiză de regresie.

Un alt aspect ce merită să fie lămurit se referă la stabilirea unei ierarhii a importanței şi a ponderii predictorilor în estimarea evoluției criteriului. În acest domeniu, coeficienții nestandardizați b îşi dovedesc limitele, deoarece sunt dependenți de amplitudinea scalelor de măsurare utilizate. De exemplu, valoarea aparent foarte mare b = 217,93 din dreptul variabilei dummy „exp_vanz" se datorează faptului că modificarea cu o unitate, de la 0 la 1, indică, de fapt, întreaga scală de răspuns, pe când un b de 24,47 indică evoluția criteriului în condițiile modificării tot cu o unitate a scalei de extroversiune, să zicem de la 0 la 1, chiar dacă răspunsurile sunt cuprinse de la 0 la 24 de puncte. Dacă b ar fi desemnat schimbarea de la 0 la 24, cu siguranță valoarea sa ar fi fost mult mai mare (de 24 de ori).

Pentru a elimina acest neajuns al coeficienților nestandardizați b, SPSS-ul oferă şi valorile coeficienților standardizați, denumiți beta. Aceştia au proprietatea de a nu fi dependenți de scala de măsurare, prin urmare ei permit o comparare directă a ponderii predictorilor. Din acelaşi ultim tabel al figurii 6.10 se constată că, în ordine descrescătoare a importanței lor în estimarea criteriului se află: un nivel ridicat de extroversiune ($\beta = .420$), experiența profesională în domeniul vânzărilor ($\beta = .368$), atractivitatea fizică ($\beta = .266$) şi un nivel scăzut de neuroticism ($\beta = -.167$). Cealaltă variabilă dummy, nefiind semnificativă statistic, nu prezintă interes, deoarece nu ajută la estimarea performanțelor în vânzări.

Unii autori precum Licht (1995), au atras atenția asupra necesității de a interpreta cu mai multă precauție aceşti coeficienți standardizați beta. Astfel, nu este indicat jocul proporției ponderilor. De exemplu, este riscant să interpretăm că un β de .42 indică o importanță aproape dublă asupra estimării performanței în vânzări în comparație cu un β de .26. De asemenea, între două valori beta apropiate, ex. β .42 şi β .36 este dificil de apreciat că prima este mai mare decât a doua, dată fiind apropierea dintre cele două care face posibilă intersecția lor în cazul precizării unui interval de încredere pentru cele două valori. Prin urmare, valorile beta trebuie interpretate într-un mod relativ. De exemplu, din valorile amintite mai sus reiese, destul de clar, că experiența profesională în domeniul vânzărilor în comparație cu absența oricărei experiențe profesionale şi nivelul de extroversiune sunt cei mai importanți predictori ai performanțelor în vânzări, cel puțin comparativ cu nivelul de neuroticism sau chiar cu gradul de atractivitate fizică.

Din punct de vedere tehnic, aceşti coeficienţi standardizaţi se pot interpreta într-un mod similar cu cei nestandardizaţi. Dacă cei din urmă arată cu cât se modifică criteriul în condiţiile în care predictorul corespunzător creşte cu o unitate, iar ceilalţi predictori rămân constanţi, coeficienţii standardizaţi arată cu cât se modifică criteriul în abateri standard dacă predictorul în cauză creşte cu o abatere standard, iar ceilalţi rămân constanţi. De multe ori, interpretarea pe baza acestor coeficienţi devine dificil de înţeles sau fără un sens foarte clar. De exemplu, în cazul extroversiunii, β de .42 arată că, în condiţiile menţinerii constante a influenţei celorlalţi predictori, atunci când nivelul de extroversiune creşte cu o abatere standard, criteriul creşte cu 0,42 abateri standard. Din aceste considerente, interpretarea datelor se face, de obicei, din prisma coeficienţilor nestandardizaţi de regresie, valorile beta servind doar la aprecierea importanţei şi ponderii predictorilor în estimarea evoluţiei criteriului.

Ce sunt coeficienţii de regresie din punct de vedere statistic?

În cazul regresiei liniare simple, beta este egal cu valoarea coeficientului de corelaţie simplă dintre predictor şi criteriu. În cazul regresiei multiple, datorită prezenţei mai multor predictori, beta nu mai este egal cu valoarea coeficientului de corelaţie simplă între predictorul respectiv şi criteriu. În astfel de cazuri, beta este un coeficient standardizat ponderat de regresie (în engleză „standardized partial regression weight"), fiind din acest punct de vedere similar cu cel prezentat în matricea modelului factorial, din cadrul analizei factoriale. În practică, se omite de multe ori sintagma „partial" din titulatură, ceea ce poate duce la interpretări eronate (Licht, 1995). Mai exact, coeficientul reprezintă o corelaţie semiparţială între predictor şi criteriu. Dacă corelaţia parţială indică relaţia dintre predictor şi criteriu, după ce se elimină influenţa celorlalţi predictori incluşi în ecuaţia asupra celor doi termeni, corelaţia semiparţială arată relaţia dintre predictorul în cauză şi acea parte a criteriului care nu a fost explicată de ceilalţi predictori din model.

6.3.2. Rolul ecuaţiei de regresie în interpretarea datelor cu scop predictiv

În urma oricărei analize de regresie efectuată în scop predictiv, ecuaţia de regresie are un rol esenţial. Prin intermediul ei, putem estima rezultatele altor persoane, pe baza predictorilor găsiţi relevanţi pentru evoluţia criteriului. Să luăm ecuaţia de regresie pentru exemplul dat, extrasă din acelaşi ultim tabel al figurii 6.10, pornind de la coeficienţii de regresie:

Performanţă vânzări = 1038 + 24,5*Extroversiune – 8,9*Neuroticism + 33,9*Atractivitate fizică + 217,9*Dummy Experienţă în vânzări + 13,6*Dummy Experienţă în alte domenii

Pe baza acestei ecuaţii, se pot estima rezultatele unui individ care, să presupunem, obţine 10 puncte la proba de extroversiune, 5 puncte la scala de

neuroticism, primeşte nota 7 ca grad de atractivitate fizică şi care are experienţă în domeniul vânzărilor. Ecuaţia de regresie în cazul acestui individ devine:

Performanţă estimată în vânzări = 1038 + 24,5*10 – 8,9*5 + 33,9*7 + 217,9*1 + 13,6*0 = 1693,7 EURO

Trebuie remarcat un alt aspect tehnic, legat de păstrarea celei de-a doua variabile dummy în ecuaţie, deşi ea nu a contribuit semnificativ statistic la estimarea valorii criteriului. Acest lucru a fost necesar, deoarece cealaltă variabilă dummy a fost semnificativă statistic. În mod normal, variabilele care nu contribuie semnificativ la estimarea criteriului sunt eliminate din ecuaţie, această măsură fiind urmată de o altă analiză de regresie, obţinută doar pe baza predictorilor semnificativi, deoarece valorile b ale acestora se modifică prin renunţarea la anumiţi predictori. Singura excepţie o constituie variabilele de tip dummy, care sunt concepute împreună ca un set de predictori. În cazul lor sunt posibile doar două decizii: fie se păstrează toate variabilele, dacă cel puţin una dintre ele este semnificativă (precum în exemplul dat), fie se elimină întreg setul de variabile dummy, în cazul în care nici una dintre acestea nu contribuie semnificativ la estimarea criteriului.

6.3.3. Instrumente complementare regresiei în scop predictiv

6.3.3.1. Decizia organizaţională

În urma stabilirii unei ecuaţii de regresie, se poate constata eficienţa modelului estimativ propus şi prin alte mijloace decât consultarea valorii coeficientului de determinare multiplă. Astfel, am putea împărţi subiecţii pe seama mediei ori a medianei, în funcţie de distribuţia statistică avută, în două grupe: subiecţii care au obţinut rezultate sub valoarea mediei (medianei), respectiv subiecţii care au obţinut rezultate peste valoarea mediei (medianei). Aceeaşi împărţire se poate face utilizând rezultatele estimate ale subiecţilor, utilizând însă drept criteriu, media sau mediana acestor estimări. În urma acestui demers, vom obţine un tabel, precum cel de mai jos, în care vor rezulta patru categorii de subiecţi: cei pentru care atât rezultatele observate, cât şi cele estimate, se află deasupra mediei (1); cei pentru care atât rezultatele observate, cât şi cele estimate se află sub nivelul mediei (4); cei ale căror rezultate observate sunt deasupra mediei însă rezultatele estimate se află sub acest indicator (fals negativ) (2); respectiv cei ale căror rezultate observate se află sub nivelul mediei, dar a căror estimare se află deasupra acestei valori (fals pozitiv) (3).

Tabelul 6.2. Distribuția subiecților pe categorii conform acestei metode pătratice

	Rezultate observate sub medie	Rezultate observate peste medie
Rezultate estimate sub medie	+ + (4)	- + (2)
Rezultate estimate peste medie	+ - (3)	- - (1)

Conform acestui tabel, cu cât procentul persoanelor ale căror rezultate au fost estimate corect (cadranele 1 şi 4) este mai mare, cu atât modelul se dovedeşte mai util în clasificarea performanței subiecților pe baza ecuației de regresie obținute.

În cazul utilizării ecuației de regresie din exemplul prezentat, pornind de la baza de date „regresie multipla.sav", sunt necesare mai multe operații.

Prima dintre ele vizează salvarea valorilor estimate ale criteriului, ca variabilă nouă, în urma rulării analizei de regresie în cauză (calea „Analyze"-Regression"-Linear"-„Save"-Unstandardized").

Cel de-al doilea pas constă în calcularea mediei şi a medianei pentru cele două variabile de interes – criteriul observat, respectiv criteriul estimat – şi alegerea celui mai potrivit indicator descriptiv pentru felul în care arată distribuția datelor. În cazul dat, distribuția pare normală pentru ambele cazuri, iar valoarea mediei a fost de 1407,27 pentru datele observate, respectiv de 1407,266 pentru datele estimate, fiind practic egale.

Al treilea pas constă în recodarea celor două variabile în două variabile diferite, nou-create, astfel încât valorile mai mari decât media să ia valoarea 1, iar cele mai mici decât media să ia valoarea 0 (calea „Transform"-„Recode"-„Into Different Variables").

În urma acestui pas, putem realiza un tabel de contingență cu rezultatele obținute:

Tabelul 6.3. Situația clasificării subiecților în exemplul dat

	Rezultate observate sub medie – 0	Rezultate observate peste medie – 1
Rezultate estimate sub medie – 0	26 (B)	6 (C)
Rezultate estimate peste medie – 1	5 (D)	23 (A)

Interpretarea acestui tabel se face simplu: din totalul de 60 de subiecți testați, pe baza ecuației de regresie 49 dintre ei au fost identificați corect. Astfel, 26 dintre cei care au obținut rezultate sub medie, au fost estimați corect, pe baza ecuației de regresie, ca obținând rezultate sub medie. Similar, în 23 de cazuri, s-a reuşit identificarea corectă a persoanelor care au obținut rezultate superioare mediei. S-au înregistrat însă şi eşecuri, la şase subiecți s-au estimat rezultate sub medie, însă ei au obținut rezultate peste medie, respectiv la cinci subiecți s-au estimat rezultate peste medie însă s-au obținut rezultate sub medie. Aceste eşecuri

sunt normale, din moment ce R^2 ajustat corespunz•tor ecuației de regresie nu a luat valoarea 1, ci doar 0,445.

Pe baza acestui tabel se poate obține şi un procentaj de răspunsuri corecte, pe baza formulei:

$$Corect = \frac{A+B}{A+B+C+D} \qquad (6.2)$$

În cazul dat, valoarea obținută indică 0,8166, ceea ce arată un procentaj de 81,66% de estimări corecte. Un asemenea rezultat se apreciază subiectiv. De exemplu, a estima corect în aproximativ 4 din 5 cazuri performanța în vânzări obținută de candidații la un post în domeniu poate fi considerată o cifră destul de bună, pentru a include predictorii respectivi în decizia procesului de selecție. În alte situații, mai ales din domeniul medical, un asemenea procentaj nu este satisfăcător, mai ales dacă el priveşte atribuirea unui diagnostic greşit, însă în situații speciale, poate fi acceptat, dacă alte metode nu reuşesc o predicție mai bună.

Pitariu (1983, p. 140-145) oferă o analiză în detaliu a acestor aspecte, oferind o serie de indicatori ce pot fi calculați şi utilizați în decizii de selecție profesională. Merită să fie menționat faptul că, în asemenea situații, împărțirea subiecților în satisfăcători şi nesatisfăcători nu trebuie să se facă ținând seama neapărat de medie. De exemplu, dacă suntem interesați de candidații foarte buni pentru un anumit post, ar fi nepotrivită împărțirea subiecților în cele două grupe pe baza mediei sau a medianei, deoarece un rezultat puțin mai mare decât aceşti indicatori nu înseamnă un rezultat foarte bun. De exemplu, dacă cei care efectuează selecția candidaților într-un post de vânzări consideră că un angajat foarte bun ar trebui să vândă lunar în valoare de 1700 de EURO, am putea utiliza acest criteriu de împărțire a candidaților în „potriviți" şi „nesatisfăcători". Utilizând acelaşi procedeu explicat mai sus, în cazul segmentării subiecților pe baza mediei, şi apelând la aceeaşi bază de date vom obține următoarea distribuție a subiecților:

Tabelul 6.4. Situația clasificării subiecților considerați „foarte buni" în exemplul dat

	Rezultate observate până în 1700	Rezultate observate peste 1700
Rezultate estimate până în 1700	45 (B)	8 (C)
Rezultate estimate peste 1700	4 (D)	3 (A)

Din acest tabel se vede că în ceea ce priveşte identificarea subiecților cu rezultate foarte bune în domeniu, lucrurile nu stau la fel de bine. Astfel, dintre cei 11 subiecți care au obținut performanțe superioare valorii de 1700 de EURO, intrând în categoria personalului din vânzări foarte bun, doar trei au fost identificați corect de către model. Ceva mai bine stau lucrurile în cazul celor considerați a nu face parte din categoria celor buni. Astfel, dintre cei 49 de participanți care au

obținut performanțe în vânzări sub valoarea de 1700 EURO, 45 au fost identificați corect. Prin urmare, modelul este destul de bun pentru a detecta pe cei care nu pot face parte din categoria subiecților „foarte buni" în domeniul vânzărilor (falşi pozitivi) însă nu este foarte bun în evaluarea subiecților care fac parte din categoria celor foarte buni, din cauza numărului relativ mare de subiecți falşi negativi. Totuşi, în ansamblu, modelul predictiv poate fi folosit în lipsa altora mai eficiente, deoarece condiția minimă de acceptare a modelului, şi anume, ca estimarea pe baza predictorilor incluşi să fie mai bună decât cea realizată în absența acestora, este îndeplinită. Formula de verificare a acestei condiții este (Pitariu, 1983):

$$\frac{A}{A+D} > \frac{A+C}{A+B+C+D} \qquad (6.3)$$

Înlocuind datele din tabelul 6.4. în formula de mai sus, obținem (3/7 > 11/60), adică 0,245, ceea ce se traduce prin faptul că, în urma utilizării ecuației de regresie respective în selecția personalului în vânzări, creşte procentul celor care vor corespunde cerințelor postului cu 24,5%. Mai multe detalii despre modul de utilizare a regresiei în procesul de selecție pot fi găsite în Pitariu (1983).

6.3.3.2. Contravalidarea

În cazul regresiei utilizate în scop predictiv, o importanță deosebită o are contravalidarea sau validarea transversală (în engleză „cross-validation"). Indiferent de termenul folosit, prin acest procedeu se urmăreşte să se stabilească în ce măsură o ecuație de regresie obținută pe baza chestionării unui eşantion de subiecți poate fi folosită cu succes în predicția performanței altor subiecți. Pentru a realiza acest tip de validare există trei mijloace de bază, primele două fiind şi cele mai potrivite. Astfel, o primă posibilitate constă în înjumătățirea eşantionului testat printr-o metodă aleatorie, urmată apoi de calcularea coeficienților ecuației de regresie pentru cele două grupe. De exemplu, dacă au fost testați 250 de subiecți, se pot obține două grupe a câte 125 de subiecți care sunt analizați prin metoda regresiei. Dacă coeficienții de regresie din cele două ecuații sunt apropiați ca valoare, se poate estima că ecuația de regresie obținută prin agregarea celor două grupe este destul de solidă şi poate fi folosită în predicția evoluției altor subiecți. Avantajul acestei metode constă în economia de timp şi resurse, iar dezavantajul constă în numărul relativ mare de subiecți ce trebuie testați pentru a putea împărți eşantionul în două grupe. De exemplu, dacă avem în total doar 100 de subiecți testați, înjumătățirea lor va conduce la un număr destul de mic de subiecți pentru a estima corect coeficienții de regresie, deoarece un eşantion mic conduce la estimarea unor coeficienți de regresie cu un grad ridicat de instabilitate. În asemenea condiții am putea ajunge la concluzia că nu vom putea utiliza ecuația pe alți subiecți, deşi instabilitatea coeficienților de regresie s-ar datora numărului mic

de subiecți testați şi nu unor caracteristici ale eşantionului care nu permit generalizarea.

O altă posibilitate constă în testarea altor subiecți şi compararea rezultatelor celor două ecuații de regresie. Această metodă este cea mai bună, din perspectiva validității ecologice însă necesită un consum mare de resurse.

În sfârşit, cea de-a treia modalitate ar fi să consultăm estimativ valoarea ajustat• a coeficientului de determinare multipl•, R^2 ajustat oferit de SPSS, care, deşi nu reprezintă ceea ce multă lume crede, şi anume, eficiența modelului de regresie în cadrul populației din care a fost extras eşantionul testat (Field, 2000), oferă o idee despre eficiența utilizării ecuației de regresie şi pe alți subiecți decât cei testați. Astfel, dacă valoarea sa ajustată este suficient de mare (ex. R^2 ajustat = .40) şi nu diferă mai mult de trei-patru procente de valoarea lui R^2 neajustat•, există premise pentru a putea generaliza modelul estimativ pe toți subiecții din care a fost extras eşantionul. Această ultimă posibilitate este în mai mare măsură supusă riscului de a ne înşela în comparație cu primele două. De aceea, recomand ca în situația utilizării regresiei în scop predictiv, să se realizeze o contravalidare, fie prin analiza dublă a grupelor extrase prin înjumătățirea lotului de subiecți, fie prin testarea unui nou eşantion şi compararea rezultatelor obținute pe cele două loturi.

Revenind la exemplul dat, numărul redus de subiecți testați (N = 60) nu ne permite aplicarea metodei de înjumătățire. Cum nu există un alt eşantion de comparație, va trebui să ne mulțumim cu aprecierea valorii ajustate a coeficientului de determinare multiplă. În condițiile în care R^2 ajustat este .445, iar R^2 este .492, se observă o distanță destul de mare între cele două valori, fapt care pune sub semnul îndoielii posibilitatea de a generaliza rezultatele obținute. Totuşi, nu ne putem baza foarte mult pe aceste rezultate, deoarece diferența destul de mare dintre cele două valori se poate datora numărului relativ redus de subiecți testați.

În ultimii ani s-au dezvoltat alte metode de contravalidare internă a modelelor de regresie precum bootstrap (o traducere aproximativă a termenului din limba engleză este de lărgire sau întindere). Bootstrap-ul (Thompson, 1995b) implică copierea setului de date de câteva sute sau mii de ori într-un fişier foarte mare. Apoi, din acesta sunt extrase sute sau mii de eşantioane diferite a căror mărime este egală cu cea a eşantionului inițial. Tehnica permite datelor deja extrase să fie selectate din nou, fiind astfel diferită de extragerile loto, în care un număr, o dată extras, nu mai poate fi pus la loc pentru a putea fi ales din nou. Pentru fiecare dintre sutele sau miile de eşantioane obținute se calculează rezultatele individuale, în funcție de scopul urmărit (ex. medii, abateri standard, coeficienți de regresie etc.). De exemplu, se poate obține o medie generală pentru coeficientul de regresie corespunzător unui anumit predictor pe baza coeficienților de regresie observați în fiecare eşantion extras. Prin aceeaşi metodă se calculează şi eroarea standard a acestui coeficient general, care nu este altceva decât abaterea standard obținută în urma împrăştierii mediilor din eşantioanele extrase. Dacă această eroare standard ia o valoare redusă, ne putem încrede în

coeficientul de regresie obținut, aspect esențial în orice analiză de regresie, cu atât mai mult în cea cu scop predictiv.

6.3.3.3. Analiza cazurilor influente

În stabilirea ecuației de regresie, nu toate cazurile, adică nu toți subiecții, contribuie la fel la stabilitatea şi corectitudinea ecuației de regresie găsite. Regresia, fiind bazată pe ideea asocierii dintre variabile, are toate limitele prezentate şi în cazul corelației. Spre exemplu, aşa cum valoarea coeficientului de corelație poate fi afectată grav de prezența unor date neobişnuite şi linia de regresie poate fi grav afectată de asemenea valori. Figura 6.11. oferă o ilustrație a acestei situații, în figura din stânga fiind prezentată linia de regresie în absența cazului neobişnuit, iar în dreapta se poate observa evoluția ei în prezența acestui caz.

Simpla prezență a unor date neobişnuite poate afecta linia de regresie şi conduce la rezultate distorsionate. Astfel, dacă analizăm linia de regresie din figura din dreapta se observă că traiectoria ei nu urmăreşte corect evoluția majorității punctelor, adică a cazurilor observate, deoarece ține seama de influența datelor extreme asupra modelului. Dacă asemenea date ar fi eliminate, precum în imaginea din stânga, linia de regresie va descrie mult mai bine evoluția criteriului. Prin urmare, în cazul utilizării regresiei în scop predictiv trebuie să analizăm neapărat dacă există date neobişnuite şi să neutralizăm influența acestora prin realizarea unei noi regresii după eliminarea lor din analiză, deoarece ele afectează valorile coeficienților de regresie.

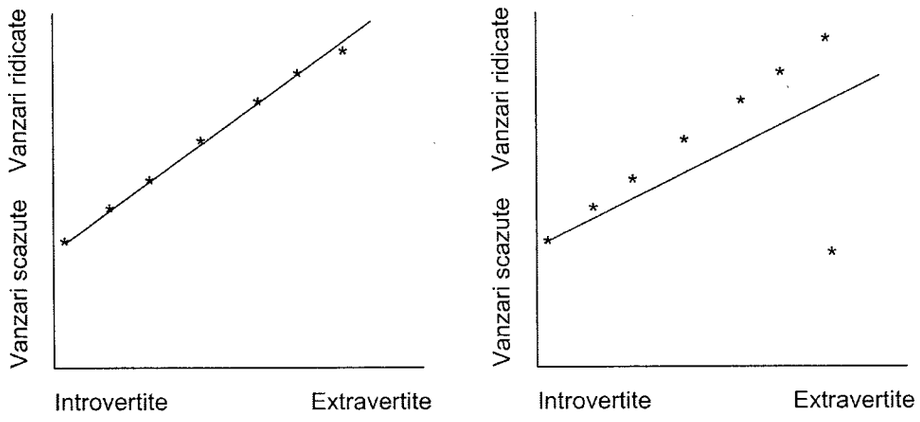


Figura 6.11. *Influența datelor neobişnuite asupra pantei liniei de regresie*

Pentru a identifica cazurile neobişnuite, în vederea obținerii coeficienților de regresie cei mai potriviți pentru a descrie setul de date, se pot realiza următorii

paşi. Se calculează reziduurile, care aşa cum am amintit, constituie diferența dintre scorul estimat şi cel observat în ceea ce priveşte criteriul. Deoarece reziduurile sunt dependente de unitatea de măsură utilizată în evaluarea criteriului, se poate calcula rezidul standardizat, în stil clasic (în engleză „standardized reziduals"), asemănător cu cel prezentat în secțiunea destinată cotelor z sau printr-un artificiu de calcul (în engleză „studentized reziduals"). Ultimul pare a fi mai eficient în estimarea corectă a valorilor standardizate în cazul reziduurilor. Interpretarea acestora se face într-un mod similar cu cel prezentat în standardizarea datelor. Astfel, în mod normal aproximativ 95% din reziduurile standardizate ar trebui să aibă valori cuprinse între -2 şi +2, iar aproximativ 99% ar trebui să aibă valori între -3 şi +3. Dacă observăm un număr mult mai mare de valori situate în afara acestor intervale, să zicem 4% se află în afara intervalului [-3 +3], am putea afirma că ecuația de regresie obținută este influențată în măsură destul de mare de prezența cazurilor extreme. Dacă acestea se află într-un număr redus (ex. 1% din reziduurile standardizate au valori cuprinse în intervalul [-3 +3]), atunci influența datelor neobişnuite asupra ecuației de regresie poate fi socotită minimă, şi în consecință, modelul estimativ poate fi utilizat în predicție.

Nu numai datele neobişnuite pot exercita o influență mare asupra coeficienților din ecuația de regresie. Alături de acestea pot fi întâlnite şi alte cazuri, denumite influente, care pot afecta valorile coeficienților de regresie. Logica acestora este simplă: dacă vom elimina un anumit caz din setul de date, iar acest lucru va conduce la modificări importante în valorile anumitor coeficienți de regresie, înseamnă că acel caz exercită o influență importantă asupra modelului estimativ. Pentru a verifica existența cazurilor influente se pot calcula o mulțime de indicatori precum: rezidul standardizat eliminat (în engleză „studentized deleted reziduals), distanța lui Cook, distanța lui Mahalanobis etc. Dintre acestea, distanța lui Cook este cel mai uşor de interpretat. Dacă valoarea sa este mai mare de 1, atunci cazul respectiv are o influență mare asupra parametrilor modelului (Cook şi Weisberg, apud Field, 2000).

Concluzia care ar trebui desprinsă din analiza diverşilor indicatori este să se păstreze, pe cât posibil, toate cazurile observate în analiza de regresie. Totuşi, în situații excepționale, în care anumite cazuri constituie atât date neobişnuite clare, cât şi cazuri care influențează în mare măsură ecuația de regresie obținută (valori ale distanței lui Cook mai mari de 1), acestea pot fi eliminate, pentru a obține un model estimativ cât mai corect. Această recomandare trebuie urmată mai ales în situația în care regresia este folosită în scop predictiv, deoarece va reduce rezidul (distanța) dintre valorile observate şi cele estimate pe baza ecuației de regresie.

În problema dată, vizând estimarea performanței subiecților în vânzare, datele indică absența unor cazuri neobişnuite şi / sau influente, aşa cum rezultă şi din outputul oferit de SPSS 11.0.

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1026,23	1814,13	1407,27	197,567	60
Std. Predicted Value	-1,929	2,059	,000	1,000	60
Standard Error of Predicted Value	41,297	147,192	64,695	14,897	60
Adjusted Predicted Value	1018,69	1813,29	1408,34	197,618	60
Residual	-381,07	545,66	,00	200,761	60
Std. Residual	**-1,816**	**2,600**	**,000**	**,957**	**60**
Stud. Residual	**-1,897**	**2,736**	**-,002**	**1,004**	**60**
Deleted Residual	-415,65	604,10	-1,07	221,466	60
Stud. Deleted Residual	-1,945	2,920	,002	1,026	60
Mahal. Distance	1,302	28,044	4,917	3,526	60
Cook's Distance	**,000**	**,134**	**,017**	**,027**	**60**
Centered Leverage Value	,022	,475	,083	,060	60

a. Dependent Variable: Vanzari lunare in EURO

Figura 6.12. *Indicatori descriptivi în analiza datelor neobişnuite şi a cazurilor influente*
SPSS este o marcă înregistrată.

Informaţiile de interes maxim au fost îngroşate. Se observă că nu există nici un caz neobişnuit (valorile reziduurilor standardizate ajustate fiind cuprinse între -1,89 şi 2,73) şi nici un caz influent (distanţele Cook fiind cuprinse între 0 şi 0,13). Dacă s-ar fi întâlnit asemenea cazuri, acestea trebuiau identificate, consultând noua variabilă creată şi denumită automat de SPSS coo_1.

6.4. Regresia multiliniară în scop explicativ

Există multe întrebări interesante la care se poate oferi un răspuns ştiinţific, fără a fi interesaţi însă de evoluţia unei anumite persoane. De pildă, am putea studia care trăsături de personalitate pot prezice adulterul, dacă consumul de tutun inhibă sau facilitează procesul de învăţare, dacă stilul de personalitate afectează performanţa în conducerea autovehiculelor etc. În toate aceste situaţii ne interesează variabilele respective, care pot oferi un răspuns epistemic deosebit de util şi nu subiecţii utilizaţi, pentru că nu ne propunem să facem anumite predicţii concrete, precum cele întâlnite în selecţia profesională. În consecinţă, dacă în regresia de tip predictiv eram interesaţi, în ultimă instanţă, de valoarea estimată a performanţei subiecţilor pe baza unei ecuaţii de regresie obţinute, în cazul regresiei cu scop explicativ ne interesează procesul în sine, fără a acorda o importanţă estimării valorilor subiecţilor.

6.4.1. Logica regresiei utilizate în scop explicativ

Regresia cu scop explicativ este utilizată frecvent atât în cercetările fundamentale, cât şi în cele aplicative. În absenţa unui design experimental care să permită atât manipularea variabilelor independente, cât şi controlul acelora care ar putea afecta relaţia dintre VI şi VD, este dificil să izolezi o relaţie de tip cauzal. Mai mult, datorită unor situaţii specifice sau a încălcării unor standarde etice de cercetare, experimentul nu poate fi utilizat, chiar dacă ne dorim să studiem o relaţie de tip cauză-efect. În asemenea situaţii, regresia multilinară reprezintă metoda de compromis, prin care se încearcă stabilirea unei relaţii nedistorsionate între variabila independentă (predictor) şi variabila dependentă (criteriu), în condiţiile controlului statistic al altor variabile de interes. De exemplu, dacă am dori să vedem care este influenţa factorilor de personalitate asupra conducerii agresive a automobilului, ar trebui să ţinem seama de o serie de aspecte intermediare care pot afecta relaţia dintre ei. Astfel, nu puţini dintre noi am văzut, de obicei seara, maşini străine conduse de „puşti" ce ascultă muzica la maxim gonind nebuneşte pe străzi, punând în pericol viaţa oamenilor şi a traficului din acele momente. Ar fi imposibil să realizăm un experiment pe această problemă numai şi din motivul că trăsăturile de personalitate nu pot fi manipulate experimental. În absenţa acestei manipulări, ne rămâne posibilitatea controlului efectelor altor variabile posibile asupra variabilei dependente. De pildă, tipul de maşină, sexul şi vârsta subiecţilor, prezenţa sau absenţa altor persoane în automobil etc. sunt numai câţiva factori care pot influenţa conducerea agresivă a maşinii. Pentru a controla aceşti factori avem două opţiuni. Fie vom exercita un control în designul cercetării, prin alegerea unui eşantion de subiecţi similari (au acelaşi gen de maşină, acelaşi număr de persoane în maşină, acelaşi sex şi aceeaşi vârstă), aspect care nu ne avantajează din perspectiva resurselor avute la dispoziţie, fie vom exercita un control statistic asupra influenţei acestor factori în relaţia dintre trăsăturile de personalitate şi conducerea agresivă a autoturismului.

Regresia liniară multiplă în scop explicativ, aplicată prin metoda ierarhică, este calea principală de control statistic al influenţei altor variabile asupra relaţiei dintre cele de interes. În cazul dat, prin regresie ierarhică putem răspunde la întrebarea: „influenţează anumite trăsături de personalitate tendinţa subiecţilor de a conduce maşina în mod agresiv, în condiţiile controlării (eliminării) influenţei altor factori situaţionali precum tipul maşinii, caracteristicile demografice ale şoferilor etc.?".

Drept exemplu, vom utiliza o bază de date ipotetică, adaptată după cercetarea realizată de Miles şi Johnson (2003). Prin comportamentul de conducere agresivă se înţelege un ansamblu de practici precum conducerea automobilului cu viteză excesivă, inadecvată situaţiei în trafic, nesocotirea unor reguli de circulaţie de bază, nervozitate la volan asociată cu adresarea de înjurături sau gesturi obscene celorlalţi participanţi la trafic, stare de surescitare sau stres intens. Aceşti autori au identificat trei dimensiuni de personalitate din Big Five care ar exercita o influenţă semnificativă asupra conducerii de tip agresiv. Astfel, o

conştiinciozitate şi o agreabilitate scăzută, asociate cu un nivel ridicat de neuroticism favorizează conduita agresivă la volan.

Dacă am utiliza regresia, în forma sa simultană, am putea vedea care este ponderea fiecăruia dintre cei trei factori, însă nu am putea controla influența altor factori (ex. vârsta, nivelul de educație, sexul persoanei etc.) asupra conduitei agresive. Prin urmare, ar exista riscul ca relațiile observate între trăsăturile de personalitate şi conducerea agresivă a maşinii să se datoreze acestor variabile terțe, pe care nu le-am controlat. De aceea, pentru a vedea în ce măsură aceste trăsături de personalitate influențează comportamentul la volan al subiecților în condițiile eliminării influenței altor factori asupra acestei conduite în trafic, va trebui să utilizăm regresia liniară ierarhică, tehnică utilizată frecvent în cazul regresiei cu scop explicativ.

Ca orice regresie de tip ierarhic, ea presupune compararea eficienței explicative a cel puțin două modele ierarhice. În cazul de față, ținând cont că variabilele de control utilizate au fost sexul persoanei, tipul de maşină (străină sau autohtonă), vârsta şi nivelul de educație (superioare, respectiv medii sau gimnaziale), cele două ecuații de regresie sunt:

Agresivitate la volan = b0 + b1*Sex + b2*Tip maşină + b3*Vârstă + b4*Educație

Agresivitate la volan = b0 + b1*Sex + b2*Tip maşină + b3*Vârstă + b4*Educație + b5*Neuroticism + b6*Agreabilitate + b7*Conştiinciozitate

Dacă diferența dintre puterea explicativă a celor două ecuații este suficient de mare, putem afirma că trăsăturile de personalitate influențează agresivitatea la volan, chiar dacă eliminăm influența altor factori precum tipul de maşină sau caracteristicile demografice ale subiecților.

Înainte de a trece la compararea celor două ecuații, pornind de la baza de date „regresie multipla auto.sav", trebuie să discutăm pe scurt ce fel de variabile sunt trecute în prima ecuație de regresie. De obicei, aici sunt trecute variabilele care se doresc a fi controlate, fie pentru că nu prezintă interes pentru obiectivul studiului (precum în situația de față, în care ne interesează doar influența factorilor de personalitate asupra conduite agresive la volan), fie pentru că au fost dovedite anterior ca având o influență semnificativă asupra variabilei dependente. O asemenea situație apare în cercetările complexe în care se doreşte izolarea unui efect în condițiile unei acțiunii comune a unei categorii de factori. De pildă, în cercetările vizând rolul nociv al fumatului asupra sănătății, acesta a putut fi susținut numai după eliminarea (controlarea) altor factori importanți pentru evoluția stării de sănătate a unei persoane precum poluarea, antecedentele familiale, vârsta, etc.

Astfel, vom include în această primă ecuație, fie factorii de interes despre care există suficiente informații că ar influența evoluția variabilei dependente, fie factorii care nu fac obiectul studiului respectiv, dar pot afecta relația dintre variabila independentă (VI) şi cea dependentă (VD). De asemenea, deşi se obişnuieşte utilizarea titulaturii de VI pentru predictor, respectiv de VD pentru criteriu, regresia

nu dovedeşte existența unei relații cauzale, aceasta fiind demonstrabilă doar prin apelul la metoda experimentului. Totuşi, pe baza unei analize logice bazate pe antecedența relațiilor de genul „fumatul afectează grav sănătatea" şi nu „sănătatea afectează grav fumatul", am putea deduce de multe ori direcția relației cauzale.

În cea de-a doua ecuație, pe lângă factorii prezentați în primul model, va fi inclus şi setul de predictori (predictorul) despre care presupunem că influențează, de asemenea, evoluția VD.

Pentru a realiza în SPSS o regresie ierarhică vom apela la calea: „Analyze"-„Regression"-„Linear". În fereastra deschisă vom introduce agresivitatea la volan ca variabilă dependentă şi predictorii din prima ecuație ca variabile independente. Apoi vom apăsa butonul „Next", iar de această dată vom introduce atât predictorii incluşi în primul pas, cât şi variabilele independente de interes, în cazul dat: „agreabil", „neurotic" şi „constiin".

6.4.2. Testarea eficienței explicative a modelului propus

Prin intermediul rezultatelor oferite de SPSS 11.0 şi prezentate în figura 6.13., vom obține răspunsul la întrebarea: contribuie trăsăturile de personalitate la explicarea conduitei agresive la volan în condițiile în care vom controla influența factorilor demografici care afectează comportamentul şoferilor?

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,562[a]	,316	,303	1,859	,316	24,926	4	216	,000
2	,688[b]	,473	,456	1,643	,157	21,202	3	213	,000

a. Predictors: (Constant), EDUCATIE, MASINA, GENDER, VARSTA

b. Predictors: (Constant), EDUCATIE, MASINA, GENDER, VARSTA, constiinciozitate, agreabilitate, neuroticism

Figura 6.13. *Compararea eficienței modelelor*
SPSS este o marcă înregistrată.

Din tabel pot fi desprinse câteva informații esențiale. De exemplu, se constată că 31,6% din evoluția dispersiei rezultatelor la agresivitatea la volan poate fi explicată de caracteristicile demografice ale şoferilor (vârstă, sex, educație) şi de caracteristicile maşinii. Acest procent creşte dacă luăm în calcul şi cele trei trăsături de personalitate, astfel încât nu mai puțin de 47,3% din dispersia criteriului agresivitate la volan poate fi pusă pe seama acțiunii comune a factorilor demografici, a tipului de maşină şi a acelor trei factori de personalitate selectați. Mai mult, diferența dintre cele două valori ale coeficientului de determinare

multiplă, ΔR^2 (delta R^2 sau R^2 schimbare) este de 15,7%. Toat• aceast• creştere a puterii explicative a modelului estimativ poate fi pusă exclusiv pe seama influenței celor trei dimensiuni de personalitate asupra agresivității la volan a şoferilor. Rezultate similare vom sesiza şi dacă vom ține seama de valorile ajustate ale coeficienților de determinare multiplă.

Se poate observa, de asemenea, că acest plus explicativ adus de introducerea dimensiunilor de personalitate în explicarea agresivității la volan aduce o schimbare semnificativă statistic, $F(3, 213) = 21{,}20$, $p < .001$, după ce modelul inițial, bazat doar pe factorii demografici şi tipul de maşină, constituia o soluție predictivă semnificativ mai bună decât cea bazată doar pe studiul mediei, $F(4, 216) = 24{,}92$, $p < .001$. În concluzie, deşi factorii demografici şi tipul de maşină contribuie semnificativ la predicția nivelului de agresivitate la volan, controlând influența acestor factori, dimensiunile de personalitate aduc un plus explicativ pentru nivelul agresivității la volan.

6.4.3. Interpretarea coeficienților de regresie în analizele cu scop explicativ

Dacă la nivel global lucrurile sunt clare, din figura 6.14 vom deduce, la nivel analitic, care dintre variabilele incluse în studiu ca predictori contribuie semnificativ statistic la explicarea nivelului de agresivitate la volan. Pentru a uşura procesul de interpretare vom oferi mai întâi modalitatea de codare a răspunsurilor utilizată în acest caz:

> Gender (1- bărbat; 0 – femeie); Maşină (1- străină; 0 – de fabricație autohtonă); Vârstă (variabilă numerică exprimată în ani); Educație (1 – studii superioare; 0 – studii medii sau gimnaziale); Neuroticism (variabilă numerică); Conştiinciozitate (variabilă numerică); Agreabilitate (variabilă numerică); Agresivitate la volan (variabilă numerică).

Pentru toate scalele numerice, un rezultat ridicat indică un nivel crescut al variabilei respective.

Se constată că programul oferă valorile coeficienților de regresie pentru ambele modele. În primul model, toți cei patru predictori: sexul persoanei, tipul de maşină, vârsta şi nivelul de educație iau valori semnificative statistic, fapt ce poate fi interpretat că fiecare dintre aceşti factori contribuie semnificativ la explicarea evoluției variabilei dependente. Fiind mai puțin interesați de estimarea agresivității la volan, ci mai mult de prezența sau absența unui efect în sine, în regresiile cu scop explicativ accentul se mută dinspre coeficienții nestandardizați, dependenți de scalele de măsurare utilizate, spre coeficienții de regresie standardizați. Astfel, în ceea ce priveşte primul model se observă că, mai ales bărbații, foarte tineri, cu un nivel de educație mai scăzut şi posesori ai unor maşini din import sunt cei mai susceptibili să conducă agresiv automobilul. Dintre predictori, se pare că tipul de maşină exercită cea mai mare influență ($\beta = .33$), în timp ce nivelul de educație contribuie cel mai puțin ($\beta = -.18$) la estimarea tipului de comportament adoptat la volan.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	8,605	,818		10,524	,000
	GENDER	1,191	,252	,268	4,730	,000
	MASINA	1,509	,256	,336	5,901	,000
	VARSTA	-,138	,031	-,265	-4,488	,000
	EDUCATIE	-,849	,272	-,184	-3,121	,002
2	(Constant)	10,991	,785		14,001	,000
	GENDER	1,285	,230	,289	5,593	,000
	MASINA	1,760	,229	,392	7,685	,000
	VARSTA	-,123	,029	-,238	-4,328	,000
	EDUCATIE	-,656	,243	-,143	-2,703	,007
	neuroticism	6,855E-03	,025	,014	,271	,787
	constiinciozitate	-,324	,054	-,305	-6,028	,000
	agreabilitate	-,215	,041	-,278	-5,314	,000

a. Dependent Variable: agresivitatea la volan

Figura 6.14. *Valorile coeficienților de regresie nestandardizați şi standardizați oferite de SPSS 11.0*

SPSS este o marcă înregistrată.

Dacă vom analiza rezultatele obținute din cea de-a doua ecuație de regresie, vom observa că singura variabilă care nu contribuie semnificativ la explicarea agresivității la volan este dimensiunea neuroticism, b = 0,0068 (în SPSS apare sub forma 6,8E-03). Toți ceilalți indicatori sunt factori care influențează semnificativ nivelul agresivității la volan. Mai mult, datorită utilizării metodei ierarhice, putem concluziona că dimensiunile de personalitate agreabilitate şi conştiinciozitate contribuie semnificativ la explicarea conduitei agresive a şoferilor la volan, în condițiile în care au fost eliminate influențele determinate de vârsta, sexul, nivelul de educație şi tipul maşinii. Cu alte cuvinte, în condițiile în care întâlnim persoane care au aceeaşi vârstă, sunt de acelaşi gen, au acelaşi nivel de educație şi au acelaşi tip de maşină, dar care diferă în privința trăsăturilor de personalitate, vor avea un nivel diferit de agresivitate la volan. Valorile negative ale coeficienților de regresie asociate acestor doi predictori semnalează că persoanele care au un nivel scăzut de conştiinciozitate şi agreabilitate vor avea tendința de a adopta un comportament mai agresiv decât persoanele care nu au aceste caracteristici psihologice, chiar în condițiile în care eliminăm (ținem constantă) influența factorilor demografici asupra acestei conduitei la volan. Din punctul de vedere al ponderii acestor efecte asupra variabilei dependente se constată că tot tipul maşinii contribuie cel mai mult la estimarea agresivității la volan ($\beta = .39$).

6.4.4. Indicatorii mărimii efectului în cazul regresiei multiple

Mărimea efectului poate fi calculată şi în cazul aplicării regresiei de orice tip. Dacă, în scop predictiv, accentul cade pe estimarea performanţelor viitoare ale unui subiect, pornind de la valorile anumitor predictori şi de la determinarea eficienţei per ansamblu a modelului propus, în utilizarea regresiei cu scop explicativ accentul cade pe izolarea efectelor produse de fiecare factor inclus în model. Astfel, avem nevoie de stabilirea efectului fiecărei variabile independente asupra celei dependente, deoarece indicatorii obişnuiţi, valoarea coeficientului de determinare multiplă, în forma ajustată sau nu, ne arată doar mărimea efectului rezultată în urma acţiunii combinate ale acestor variabile.

Pentru aceasta, dacă vom bifa opţiunea corespunzătoare din cadrul regresiei, accesibilă prin acţionarea butonului „Statistics", SPSS-ul ne oferă indicatorii de corelaţie parţială şi semiparţială, pe baza cărora putem stabili coeficienţi de determinare individuali, în cazul fiecărui predictor în parte.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	8,605	,818		10,524	,000			
	GENDER	1,191	,252	,268	4,730	,000	,292	,306	,266
	MASINA	1,509	,256	,336	5,901	,000	,325	,373	,332
	VARSTA	-,138	,031	-,265	-4,488	,000	-,272	-,292	-,253
	EDUCATIE	-,849	,272	-,184	-3,121	,002	-,304	-,208	-,176
2	(Constant)	10,991	,785		14,001	,000			
	GENDER	1,285	,230	,289	5,593	,000	,292	,358	,278
	MASINA	1,760	,229	,392	7,685	,000	,325	,466	,382
	VARSTA	-,123	,029	-,238	-4,328	,000	-,272	-,284	-,215
	EDUCATIE	-,656	,243	-,143	-2,703	,007	-,304	-,182	-,134
	neuroticism	6,855E-03	,025	,014	,271	,787	-,004	,019	,013
	constiinciozitate	-,324	,054	-,305	-6,028	,000	-,326	-,382	-,300
	agreabilitate	-,215	,041	-,278	-5,314	,000	-,193	-,342	-,264

a. Dependent Variable: agresivitatea la volan

Figura 6.15. *Valori ale coeficienţilor de corelaţie oferite de SPSS 11.0*
SPSS este o marcă înregistrată.

Coeficienţii de ordinul zero, respectiv coeficienţii de corelaţie parţială, nu sunt altceva decât coeficienţii de corelaţie r Bravais-Pearson, respectiv coeficienţii de corelaţie parţială discutaţi pe îndelete în capitolul al treilea, destinat tehnicilor introductive de testare a ipotezelor. Ultima coloană, denumită „part", indică valoarea coeficienţilor de corelaţie semiparţială între factorul respectiv şi variabila dependentă în condiţiile eliminării influenţei celorlalţi predictori asupra VD. De exemplu, valoarea lui r semiparţial notat r_{sp} este - .30, ceea ce înseamnă că dacă vom elimina influenţa celorlalţi predictori asupra agresivităţii la volan, există o asociere unică, inversă, între nivelul de conştiinciozitate şi agresivitatea la volan, o conştiinciozitate scăzută ducând la un nivel ridicat de agresivitate în conducerea

automobilului. Fiind vorba de un efect izolat şi totodată de un coeficient de corelaţie, se poate calcula coeficientul de determinare specific pentru relaţia dintre cele două variabile, reamintindu-ne faptul că acesta este egal cu valoarea pătratului coeficientului de corelaţie. Astfel, am putea afirma că 9% ($-.30^2$) din dispersia comportamentul agresiv la volan poate fi explicată de diferenţele interindividuale în ceea ce priveşte nivelul de conştiinciozitate al subiecţilor testaţi.

Elementele prezentate mai sus constituie cele mai importante aspecte ale interpretării statistice în cazul utilizării regresiei în scop explicativ. Alături de aceste elemente, pot fi analizate şi alte aspecte, amintite în cazul regresiei în scop predictiv, precum contravalidarea sau analiza cazurilor influente. Însă, dacă în cazul regresiei cu scop predictiv soluţia nu poate fi concepută fără o analiză atentă a acestor aspecte, deoarece ecuaţia va fi utilizată în scop estimativ, predictiv, în regresia utilizată în scop explicativ acestea sunt importante, dar nu esenţiale. Rezultatele obţinute într-o regresie cu scop explicativ pot fi reţinute şi apreciate în felul lor, fără a realiza neapărat o analiză a cazurilor influente sau fără a contravalida modelul, dacă vom respecta însă condiţiile de aplicare a regresiei. Numai îndeplinind aceste condiţii, vom avea suportul că modelul explicativ rezultat nu suferă de distorsiuni grave.

6.5. Condiţii necesare pentru aplicarea regresiei liniare

În vederea aplicării corecte a regresiei liniare este necesară îndeplinirea mai multor condiţii tehnice. O bună trecere în revistă a acestora este oferită în Berry şi Feldman (1985) sau Licht (1995).

6.5.1. Evitarea erorii de specificare

Sub această denumire pretenţioasă se regăsesc acele modele incorect alese. O asemenea situaţie apare atunci când între predictorii unei ecuaţii de regresie nu se găsesc unii care ar fi trebuit incluşi. De exemplu, dacă atât teoria, cât şi practica din domeniul vânzărilor indicau asertivitatea ca o trăsătură de personalitate de bază în estimarea performanţelor în vânzări, dar noi nu o includem în modelul predictiv estimat, comitem o eroare de specificare. Similar, în cazul conducerii agresive a automobilului, am realiza o eroare de specificare, dacă teoria din domeniu include reacţia autorităţilor faţă de acest comportament printre predictorii care determină nivelul de agresivitate la volan, iar noi nu ţinem seama de el, neincluzându-l în modelul explicativ propus. Astfel, neintroducerea tuturor factorilor relevanţi pentru problema în cauză, afectează rezultatele obţinute, existând chiar riscul de distorsiune a concluziilor, deoarece nu ţinem seama de influenţa acestor factori când ar trebui să-i luăm în considerare.

În direcție opusă, eroarea de specificare poate apărea prin includerea în model a unor predictori care nu ar fi trebuit introduşi, deoarece ei nu relaționează cu criteriul. Totuşi, această situație este mai puțin gravă decât prima, deoarece ea poate fi remediată prin eliminarea predictorului care nu contribuie semnificativ la estimarea evoluției criteriului, urmată de o nouă analizare a modelului predictiv, fără variabila în cauză. În sfârşit, un alt caz al erorii de specificare este cel în care, deşi au fost selectați predictorii relevanți, modelul unei regresii liniare nu este cel adecvat. Modelul regresiei liniare porneşte de la premisa unei relații liniare între predictori şi criteriu şi de la un model aditiv al efectului pe care-l au aceşti predictori asupra modelului de estimat. În consecință, dacă utilizăm regresia liniară, în condițiile în care nu există o relație de liniaritate între fiecare predictor şi criteriu, respectiv în cazul în care există o situație de multicoliniaritate între predictori, datorită unei corelații ridicate între aceştia, vorbim tot de o eroare de specificare.

În plan tehnic, situația unei erori de specificare conduce la obținerea unor coeficienți de regresie distorsionați sau instabili. De asemenea un alt indiciu constă într-o valoare scăzută a coeficientului de determinare multiplă (R^2). Totuşi, nu orice R^2 scăzut indică o problemă de eroare de specificare, deoarece o asemenea valoare se poate datora unui alt factor, precum includerea în model a unor predictori cu o consistență internă scăzută (implicit o eroare de măsurare ridicată), fapt des întâlnit în psihologie, prin utilizarea scalelor ca modalități de evaluare a diferitelor variabile. De aceea, cazul erorii de specificare datorat neincluderii unor variante relevante poate fi detectat doar pe baza unei teorii solide existente cu privire la evoluția criteriului. În schimb, cazul includerii unor variabile irelevante este mult mai uşor de detectat prin consultarea valorilor t corespunzătoare fiecărui coeficient de regresie. Dacă această valoare nu este semnificativă statistic, înseamnă că variabila respectivă nu contribuie la estimarea evoluției criteriului. O altă modalitate de evaluare este aceea de a compara valorile ajustate ale lui R^2, iar dacă nu se constată diferențe între modelul care include şi cel care exclude variabila în cauză, înseamnă că aceasta poate fi eliminată din modelul predictiv.

6.5.2. Evitarea erorii de măsurare

Din păcate, în psihologie, ca şi în alte domenii alte ştiințelor sociale, nu putem măsura direct constructul dorit, ci trebuie să apelăm la o serie de indicatori. De exemplu, utilizăm un test de inteligență pentru a măsura nivelul intelectual sau utilizăm o scală de agresivitate pentru a măsura magnitudinea acestei trăsături. Nu întotdeauna există o corespondență între valoarea indicatorului şi valoarea reală (latentă) a constructului teoretic respectiv. Acea diferență poate fi denumită eroare de măsurare. Ea este de două feluri: sistematică şi aleatoare. Primul caz apare în situația în care instrumentul utilizat pentru a măsura acel concept măsoară şi altceva, apărând o problemă de validitate de construct. O asemenea situație este ținută sub control prin studiile de validare a instrumentelor utilizate, de aceea, în situația utilizării unor instrumente adecvate şi validate, această eroare este

eliminată sau redusă foarte mult. În schimb, eroarea de măsurare aleatorie ține de caracteristicele psihometrice ale instrumentului utilizat, în special de fidelitatea probei, dar şi de unele deficiențe în codarea sau tehnoredactarea bazei de date.

Efectul acestui tip de eroare asupra coeficienților de regresie este diferit, dacă eroarea apare în cazul criteriului sau în cazul predictorilor. Dacă criteriul utilizat are o fidelitate redusă, scade corelația acestuia cu predictorii, fapt ce afectează atât valorile coeficienților de regresie, cât şi valoarea coeficientului de determinare multiplă, afectând astfel eficiența predictivă a modelului. Prin urmare, criteriul utilizat ar trebui să aibă o consistență internă cât mai ridicată, dacă nu chiar 1. De aceea, recomand evitarea folosirii pe post de criteriu a acelor variabile care au o consistență internă sub .80, în cazul utilizării regresiei în scop predictiv, respectiv de .70, în cazul utilizării regresiei în scop explicativ. În cazul exemplului dat, criteriul ales, performanța obținută în vânzări, nu ridică probleme, deoarece eroarea de măsurare aleatorie este absentă, singurele probleme putând apărea din deficiențe în înregistrarea corectă a datelor.

Dacă eroarea de măsurare apare în cazul predictorilor, efectul este ceva mai puțin grav, dar prezent. Într-o asemenea situație, dacă variabila în cauză nu corelează cu ceilalți predictori, va fi afectată doar valoarea coeficientului de regresie corespunzător variabilei respective însă dacă variabila corelează şi cu ceilalți predictori, vor fi afectați, într-o anumită măsură, toți coeficienții de regresie. Modul în care sunt afectați aceşti coeficienți nu este foarte clar, de aceea se recomandă utilizarea unor instrumente cu calități psihometrice cât mai bune.

6.5.3. Eroarea reziduală cât mai mică

Prin termenul reziduuri se defineşte diferența dintre valoarea observată a criteriului şi cea estimată prin utilizarea ecuației de regresie. Aşa cum am precizat, cu cât această diferență este mai mică, cu atât modelul este mai bun în estimarea evoluției criteriului. Totuşi, datorită unor aspecte matematice care nu vor fi discutate aici este de dorit ca aceste reziduuri: (1) să aibă media egală cu zero; (2) să îndeplinească condiția de homoscedasticitate, având aceeaşi dispersie pe întreaga amplitudine a predictorilor; (3) să nu coreleze cu alte reziduuri sau cu predictorii; (4) să aibă o distribuție normală (Licht, 1995). Se pare că încălcări moderate ale acestor condiții nu influențează foarte mult eficiența modelului predictiv.

6.5.4. Evitarea multicoliniarității

Sub această denumire este specificată încălcarea condiției ca predictorii aleşi pentru estimarea evoluției criteriului să nu coreleze prea puternic între ei. Prin urmare, nu se poate pune problema prezenței sau absenței multicoliniarității, ci

doar a prezenței ei într-un grad mai mare sau mai mic, în funcție de corelația existentă între predictori.

Efectul unei multicoliniarități crescute constă în creşterea probabilității de a obține coeficienți de regresie nesemnificativi statistic pentru una dintre variabilele care corelează puternic, deoarece această problemă afectează valorile t asociate acestor coeficienți, din cauza dificultății de a decela efectele fiecărei variabile asupra criteriului. Din această perspectivă, creşte riscul de a comite eroarea de tip II. Indiferent de scopul utilizării regresiei, problema unei multicoliniarități ridicate este foarte gravă. Astfel, în cazul utilizării regresiei în scop predictiv, problema principală constă în instabilitatea coeficienților de regresie, fapt ce conduce la dificultatea de a generaliza coeficienții respectivi şi de a-i utiliza pe un alt lot de subiecți. În schimb, în cazul utilizării regresiei în scop explicativ, problema capitală constă în dificultatea de a decela efectul fiecărui predictor implicat asupra evoluției criteriului (variabilei dependente). Detectarea unei situații de **multicoliniaritate** se poate face simplu prin realizarea unei corelații între fiecare pereche de predictori incluşi în studiu. Dacă există vreo corelație foarte ridicată, să spunem, peste .70, am putea intui o problemă de multicoliniaritate ce trebuie rezolvată. SPSS-ul oferă alte două modalități de diagnosticare a unei multicoliniarități ridicate prin indicatorii VIF (în engleză „variance inflation factor"), respectiv toleranța (1/VIF). Dacă valorile VIF sunt peste 10, iar cele ale toleranței sunt sub 0,10, avem de-a face cu o problemă gravă de încălcare a condițiilor de utilizare a regresiei, datorită unei multicoliniarități ridicate. Unii autori propun chiar praguri mai joase, astfel încât o valoare VIF de 5 sau o toleranță sub 0,20, poate indica probleme în estimarea corectă a modelului de regresie datorită multicoliniarității (Field, 2000).

Efectul prezenței unei multicoliniarități ridicate este vizibil în exemplul următor, al unei regresii în scop explicativ, în care avem trei predictori. Să luăm ca valori de exemplu β de .80, .30 şi .30. La o primă vedere am putea afirma că primul predictor are o influență mai mare asupra criteriului decât ceilalți doi. Totuşi, această situație se poate datora unei corelații puternice între ultimii doi predictori. Astfel, dacă am elimina unul dintre ei, valorile β pentru cei doi predictori rămaşi ar fi .70 şi .66. În asemenea condiții nu ar mai fi foarte clar care dintre predictorii luați în considerare are o influență mai mare în explicarea evoluției criteriului. Acest exemplu vine în sprijinul ideii de a interpreta în termeni relativi diferențele observate între coeficienții standardizați de regresie, deoarece aceştia pot suferi în situații de multicoliniaritate.

Pentru a elimina situația de multicoliniaritate sunt disponibile câteva soluții. Una dintre ele vizează creşterea numărului de participanți în studiu, cunoscându-se tendința ca valoarea coeficientului de corelație dintre două variabile să scadă o dată cu creşterea volumului eşantionului. O altă posibilitate constă în combinarea variabilelor care corelează puternic, fie prin alcătuirea unui scor compozit, fie prin reducerea datelor ca urmare a analizei componentelor principale. Prin această ultimă soluție se va pierde însă posibilitatea de a desprinde efectele separate ale variabilelor care corelează puternic însă se câştigă timp şi bani, nefiind nevoie de

testarea mai multor subiecți. În fine, alți autori recomandă eliminarea uneia dintre variabilele care corelează prea puternic cu alte variabile.

SPSS-ul oferă posibilitatea unui diagnostic cu privire la nivelul de multicoliniaritate prin bifarea opțiunii în cauză disponibilă prin accesarea butonului „Statistics", urmând calea „Analyze"-„Regression"-„Linear".

6.5.5. Existența unor relații de tip liniar şi aditiv

Prin liniaritatea unei relații se înțelege ideea că între fiecare predictor şi criteriu există o relație de tip liniar, adică valoarea coeficientului de regresie se menține constantă la orice nivel al predictorului. Astfel, dacă b pentru extroversiune este 24,47, performanța unui individ în vânzări creşte cu 24,47 EURO în situația în care el a obținut 3 puncte la scala de extroversiune față de unul care a obținut 2 puncte la această probă. Această diferență, de 24,47 EURO, se menține şi în compararea unei persoane care a obținut, să zicem 21 de puncte, față de o alta care a obținut 20 de puncte. Mai multe informații despre ce înseamnă o relație liniară pot fi citite în secțiunea 1.3.2. din capitolul introductiv, când am tratat problema reprezentării grafice a norului de puncte.

Prin relația aditivă se înțelege ideea că nu există un efect de interacțiune între doi predictori, cu alte cuvinte, că efectul unui predictor asupra criteriului nu este dependent de valoarea unui alt predictor. Dacă acest lucru se întâmplă, este încălcată ideea de aditivitate şi trebuie să se țină seama de interacțiunea dintre cei doi predictori, ca efect suplimentar în estimarea evoluției criteriului.

Pentru detectarea situațiilor de relații neliniare sau non-aditive cea mai bună soluție este de ordin teoretic. Dacă ne aşteptăm ca relația dintre anumiți predictori şi criteriu să fie neliniară sau să existe o interacțiune între predictori care poate explica evoluția criteriului, este bine să se specifice aceste lucruri încă din formularea modelului predictiv.

Să presupunem că avem doi predictori într-o ecuație de regresie. Dacă ne aşteptăm ca ecuația să îndeplinească ambele cerințe, atunci ea are forma simplă de:

$$Criteriu = b0 + b1*Criteriu\_1 + b2*Criteriu\_2 \quad (1)$$

Dacă ne aşteptăm ca ecuația să nu fie liniară, ci curbilinie, în cazul predictorului 1, dar liniară în cazul predictorului 2, atunci ea are forma de:

$$Criteriu = b0 + b11*Criteriu\_1 + b12*Criteriu\_1^2 + b2*Criteriu\_2 \quad (2)$$

Dacă ne aşteptăm ca ecuația să fie liniară, dar nu şi aditivă, atunci ea are forma de:

$$Criteriu = b0 + b1*Criteriu\_1 + b2*Criteriu\_2 + b12*Criteriul\_1*Criteriul\_2 \quad (3)$$

Prin urmare, efectul de aditivitate poate fi observat prin includerea unei variabile suplimentare care se obține prin produsul celor două variabile aflate în situație de interacțiune (mai multe detalii sunt prezentate în secțiunea 6.7.2., în timp ce relația de non-liniaritate poate fi tratată printr-o ecuație liniară, după includerea unei variabile reprezentate de pătratul predictorului care încalcă relația de liniaritate cu criteriul.

Detectarea statistică a unor asemenea relații, atunci când teoria nu ne ajută la estimarea tipului de relație între variabile, se poate face printr-o metodă elegantă, dar costisitoare ca resurse de subiecți. De exemplu, pentru a detecta situația de non-liniaritate se poate împărți lotul de subiecți în trei sau patru categorii, pe baza rezultatelor obținute la criteriu. Spre exemplu, vom alege grupa subiecților cu performanțe superioare, cu performanțe medii, respectiv cu performanțe inferioare în vânzări. Dacă atât interceptul, cât şi coeficienții de regresie rezultați nu diferă foarte mult în cele trei cazuri înseamnă că există o relație liniară între ansamblul de predictori şi criteriu. În schimb, dacă sunt diferențe foarte mari între aceştia (ex. b pentru extroversiune variază între 2,33 şi -4,11), atunci am putea sesiza o relație clar neliniară. În SPSS, acest lucru poate fi testat mai simplu, cerând calculatorului să ne prezinte valoarea coeficientului de determinare multiplă în situația unei relații liniare, respectiv a unei relații neliniare. În cazul în care R^2 asociat modelului neliniar este semnificativ mai mare decât cel asociat modelului liniar putem presupune existența unei relații neliniare.

Pentru detectarea unei situații de interacțiune poate fi utilizată metoda regresiei ierarhice; astfel, dacă diferențele dintre cele două valori F asociate modelelor 1 şi 3 prezentate mai sus sunt semnificative statistic, putem trage concluzia existenței unui efect de interacțiune între cei doi predictori care influențează evoluția criteriului.

6.6. Exemplu de regresie multiplă liniară – cazul citirii

Într-un studiu realizat de Kirby, Parrilla şi Pfeiffer (2003) s-a urmărit identificarea influenței unor structuri cognitive asupra deprinderii cititului. Mai concret, s-a studiat rolul pe care-l are viteza de pronunțare şi abilitatea de a emite şi manipula sunete în performanța de citire a unui text. Studiul nu are doar valențe de cercetare fundamentală, ci urmăreşte un aspect aplicativ esențial: identificarea, încă din grădiniță, a copiilor susceptibili să sufere de dislexie (dificultăți de citire). Autorii au studiat aceste efecte pe parcursul mai multor ani. Noi ne vom rezuma la rezultatele obținute de copii la proba de citire în clasa a II-a şi la o singură variabilă dependentă – capacitatea de recunoaştere a cuvintelor la testul standard Gates-MacGintie Reading Comprehension Test.

Pentru a izola influența celor doi factori a fost necesar controlul altor câtorva variabile, ce pot afecta capacitatea de a citi. Variabilele controlate au fost: nivelul de dezvoltare intelectuală (apreciat atât prin probe verbale, cât şi de

performanță); şi nivelul de familiarizare cu un material tipărit (câte litere cunoaşte copilul deja, aspect ce poate ascunde diferențe socio-demografice).

Pentru a măsura cei doi factori: viteza de pronunțare, respectiv abilitatea fonetică (abilitatea de a emite şi manipula sunete) s-au utilizat mai multe probe, care au fost supuse ulterior unei analize a componentelor principale, rezultând în final două scoruri, câte unul pentru fiecare predictor. Pentru a facilita interpretarea datelor, trebuie să precizăm că rezultatele ridicate indică performanțe superioare.

În acest exemplu, ca, de altfel, în orice ecuație de regresie utilizată în scop predictiv, se precizează câteva elemente esențiale în secțiunea de analiză a datelor:

- precizarea într-un paragraf a tipului de analiză întreprinsă (ex. regresie multiliniară ierarhică), alături de justificarea alegerii făcute; în plus, se poate face o precizare scurtă legată de respectarea sau nerespectarea condițiilor de bază necesare pentru aplicarea acestei tehnici;
- precizarea, într-un alt paragraf, a predictorilor, a criteriului şi a modului în care au fost introduşi aceştia;
- prezentarea unui tabel sumar a indicatorilor descriptivi şi a matricei de corelații între variabilele incluse în model, conform standardelor APA, după modelul oferit de Nicol şi Pexman (1999);
- prezentarea într-un tabel sumar a coeficienților standardizați de regresie pentru fiecare predictor în parte, cât şi a valorilor coeficienților de determinare multiplă aferenți modelului estimativ propus; acest lucru se poate face utilizând aceleaşi recomandări APA prezentate în Nicol şi Pexman (1999), cu precizarea că, în cazul regresiilor în scop predictiv, trebuie precizate, suplimentar, valorile coeficienților nestandardizați de regresie şi eroarea standard de măsurare a acestora;
- interpretarea pe scurt a rezultatelor obținute (o analiză cantitativă).

În cazul exemplului de față, textul din secțiunea de analiză a datelor ar putea lua forma următoare:

Tabel cu indicatorii descriptivi şi matricea de corelații între variabilele implicate în model (clasa a II-a)

Variabila	M	AS	1	2	3	4	5
Performanța de recunoaştere a cuvintelor	39.90	18.70	.16	.11	.45 **	.31 **	- .20 *
PREDICTORI							
1. Nivel intelectual – probe verbale	12.90	2.88	--				
2. Nivel intelectual – probe de performanță	6.40	2.21	.57 **	--			
3. Familiarizarea cu materiale tipărite	35.20	17.90	.49 **	.21 *	--		
4. Abilitatea fonetică	6.50	5.69	.18 *	.23 *	.36 **	--	
5. Viteza de pronunțare	42.20	11.53	.06	.20 *	.15	- .02	--

* $p < .05$; ** $p < .01$

Analiza de regresie ierarhică a fost utilizată pentru a observa influența vitezei de pronunțare şi a abilității de manipulare a sunetelor asupra performanțelor în citire obținute de copii în clasa a II-a. Am decis să utilizăm regresia ierarhică pentru a controla influența abilității intelectuale generale şi a nivelului de familiarizare cu materiale tipărite asupra performanței în citire. Merită remarcat faptul că toți predictorii au fost evaluați când subiecții se aflau încă la grădiniță, pentru a putea vedea în ce măsură pot fi anticipate dificultățile de citire, pentru a se interveni încă de timpuriu. Din puncte de vedere tehnic, nu au existat situații speciale, toate condițiile de aplicare a regresiei fiind susținute.

(Comentariu: prin acest design, cercetarea dobândeşte un caracter mixt, atât explicativ – în izolarea şi identificarea influenței celor doi factori asupra abilității de a citi, cât şi predictiv, prin posibilitatea de a estima evoluția şcolară a elevilor în ceea ce priveşte cititul, pe baza unor predictori testați anterior. Datele prezentate în primul tabel sunt fictive, deoarece autorii nu au inclus un asemenea tabel, în schimb cele prezentate în tabelul de mai jos sunt reale, dar adaptate conform cerințelor APA).

Tabel sumar cu rezultatele analizei de regresie ierarhice vizând estimarea performanțelor de recunoaştere a cuvintelor pe baza predictorilor luați în studiu (N = 161).

VARIABILE	R^2	ΔR^2	β	B	SE b
PASUL 1	.378 **	.378 **			
Abilitate intelectuală generală					
Probe verbale			.178 *		
Probe de performanță			.142		
Familiaritate cu materiale tipărite			.421 **		
PASUL 2	.479 **	.101 **			
Abilitate intelectuală generală			.		
Probe verbale			.157		
Probe de performanță			.153		
Familiaritate cu materiale tipărite			.458 **		
Abilitate fonetică			.302 **		
Viteză de pronunțare			- .209 *		

* p < .05; ** p < .01

În fiecare model ierarhic, variabila dependentă era constituită din performanța obținută în citire privind recunoaşterea literelor de către elevii din clasa a II-a. Predictorii incluşi în primul pas au fost abilitatea intelectuală generală şi gradul de familiaritate cu materialul tipărit, pentru a putea controla efectul acestora asupra variabilei dependente. În cel de-al doilea pas, pe lângă aceşti predictori au fost introduşi cei doi factori de interes, şi anume, viteza de pronunțare şi abilitatea de manipulare a sunetelor.

(Comentariu: în cazul regresiei utilizate în scop predictiv este necesară şi prezentarea coeficienților nestandardizați de regresie (b), cât şi a erorii standard de măsurare a acestora (SE b). În cazul regresiei cu scop explicativ, predominantă în articolul

de față, nu este necesară precizarea acestor indicatori suplimentari. De altfel, autorii studiului de față nici nu oferă valorile acestora, de aceea am lăsat spațiile respective necompletate).

Rezultatele din tabelul de mai sus sprijină ipoteza conform căreia atât abilitatea fonetică, cât şi viteza de pronunțare a sunetelor influențează semnificativ performanța elevilor de a recunoaşte şi citi cuvinte, în condițiile în care controlăm influența datorată diferențelor de nivelul intelectual şi de familiarizare cu materialele tipărite ($\Delta R^2 = .101$, $p < .01$). Astfel, cei doi factori explică împreună 10,1% din performanța obținută la proba de citire de către elevi, iar aceasta vine în plus, față de cele 37,8% explicate de factorii controlați (nivelul intelectual şi gradul de familiarizare cu materialele tipărite).

(Comentariu: p nu este un atribut al lui R^2, ci este caracteristic lui F schimbare. Totuşi, se obişnuieşte trecerea acestei valori în dreptul lui R^2 deoarece acesta oferă o informație în plus: ce procentaj din dispersia criteriului poate fi explicată de variația predictorilor incluşi în studiu?).

La nivel analitic, din modelul estimativ final se observă că nivelul intelectual nu influențează abilitatea de a identifica cuvinte, valorile beta nefiind semnificative statistic. Acest lucru este aparent surprinzător, însă dacă ținem seama că proba de identificare a cuvintelor are loc, în clasa a II-a, într-o şcoală normală, şi că ea nu constă în elemente de comprehensiune a textului, ci doar în elemente de citire a cuvintelor, rezultatele nu mai sunt surprinzătoare.

În schimb, se constată că elevii care sunt familiarizați încă din grădiniță cu materialele tipărite obțin rezultate superioare la testul de identificare a cuvintelor ($\beta = .45$, $p < .01$), acelaşi fapt petrecându-se şi în cazul copiilor care dispun de o abilitate fonetică superioară ($\beta = .30$, $p < .01$). În schimb, se constată o relație inversă între viteza de pronunțare şi capacitatea de a identifica cuvintele ($\beta = - .20$, $p < .01$). Acest fapt se explică prin tendința subiecților cu o viteză superioară de pronunțare de a se grăbi şi de a greşi în identificarea corectă a cuvintelor (probabil citind primele litere şi apoi ghicind cuvântul). Analiza erorilor ar putea oferi suport acestei posibile explicații.

În concluzie, datele obținute sprijină ideea că dincolo de diferențele intelectuale sau educaționale (familiarizarea cu materiale tipărite), există anumite structuri psihologice, precum abilitatea fonologică şi viteza de pronunțare, care influențează semnificativ performanțele subiecților în citire. Acest rezultat oferă psihologilor posibilitatea de predicție şi intervenție timpurie (încă din grădiniță) asupra copiilor care pot întâmpina dificultăți în citire pe perioada şcolii, realizându-se astfel primii paşi în vederea ameliorării acestei situații.

În spatele acestei secțiuni stau însă o serie de operații metodologice şi statistice precum:

- identificarea scopului regresiei: de tip predictiv, explicativ sau mixt;

- alegerea celei mai bune modalități de regresie: simultană, ierarhică etc;
- selectarea predictorilor relevanți pentru estimarea evoluției criteriului şi, dacă este cazul, a ordinii de introducere a acestora în model;
- interpretarea per ansamblu a eficienței modelului;
- interpretarea coeficienților de regresie din perspectiva ipotezelor formulate;
- verificarea îndeplinirii condițiilor de aplicare a regresiei pentru a ne asigura că valorile găsite nu sunt distorsionate grav.

6.7. Direcții de evoluție ale regresiei

6.7.1. Alte tipuri de regresie

Deşi este aplicată foarte des, regresia liniară aşa cum a fost prezentată în acest capitol, poate fi aplicată doar în situația în care criteriul (variabila dependentă) este măsurat printr-o scală numerică. Unii autori aplică regresia liniară şi în condițiile în care răspunsurile sunt evaluate pe o scală de tip Likert însă aceasta trebuie să posede minimum 7 trepte de răspuns (Garson, 2000).

În condițiile în care răspunsurile ordinale au mai puțin de şapte trepte sau în cazul în care criteriul este o variabilă de tip nominal, se aplică alte tipuri de regresie. O clasificare a acestor tipuri, în funcție de tipul de predictori şi criterii permise, este oferită în tabelul 6.5, iar una dintre cele mai întâlnite forme, regresia logistică pentru date dihotomice, va fi detaliată în capitolul următor.

Tabelul 6.5. O clasificare non-exhaustivă a tipurilor de regresie

Denumire	Predictori	Criteriu
Regresia logistică (binară)	Orice variabile, indiferent de scala de măsurare	Numai variabile dihotomice
Regresia logit	Numai variabile de tip nominal	Numai variabile dihotomice
Regresia logistică multinomială	Orice variabile, indiferent de scala de măsurare	Numai variabile nominale (exclus cele dihotomice)
Regresia logistică ordinală	Orice variabile, indiferent de scala de măsurare	Numai variabile ordinale (ex. scale Likert cu 5 clase)

6.7.2. Identificarea efectelor de interacțiune în regresia multiplă

Din majoritatea modelelor discutate în acest capitol, analiza de regresie viza identificarea efectelor pe care fiecare predictor la are asupra criteriului. Aceste

efecte izolate pot fi denumite efecte simple, fiind analoge efectelor simple din tehnicile ANOVA.

Astfel, dacă într-un design ANOVA factorial de tip 2x3, rezultau trei valori F (FA, FB, FAB), într-o regresie de tip clasic, factorul interacțiune nu poate fi detectat, deoarece ecuația de regresie, în formă standardizată este:

$$y = \beta_{A}*X_{A} + \beta_{B}*X_{B} \qquad (6.4)$$

Pentru a putea măsura şi cel de-al treilea efect, cel de interacțiune, ecuația de regresie ce trebuie testată devine:

$$y = \beta_{A}*X_{A} + \beta_{B}*X_{B} + \beta_{AB}*X_{A}*X_{B} \qquad (6.5)$$

Interpretarea primară a datelor este simplă: dacă coeficienții de regresie sunt semnificativi statistic înseamnă că predictorii respectivi contribuie semnificativ statistic la evoluția criteriului. Deşi lucrurile par simple, la prima vedere măsurarea efectului de interacțiune ridică câteva probleme ce vor fi discutate, pe scurt, în continuare. Cum obținem variabila interacțiune $X_{A}*X_{B}$? Când se poate măsura acest efect de interacțiune? Cum se poate cunoaşte dacă există un efect de interacțiune? Cum se interpretează coeficienții corespunzători unui efect de interacțiune? Cum se aplică acest efect de interacțiune în cazul variabilelor dummy?

Cum obținem variabila de interacțiune $X_{A}*X_{B}$?

În mare, există două modalități de lucru, prima fiind mai intuitivă şi mai des întâlnită, însă cea de-a doua fiind mai potrivită din punctul de vedere al caracteristicilor statistice pentru variabilele de tip interval, atât de frecvent întâlnite în domeniul psihologiei.

Prima modalitate constă în realizarea produsului termenilor celor două variabile despre care se crede că se află în relație de interacțiune (una dintre ele moderând relația celeilalte cu predictorul). Această modalitate simplă se aplică atunci când una dintre variabilele implicate este de tip dummy şi, de asemenea, poate fi aplicată fără probleme, dacă ambele variabile sunt măsurate prin scale de proporții. Dacă nu ne aflăm în vreuna din cele două situații, se preferă utilizarea celeilalte metode, deoarece simpla înmulțire a variabilelor poate conduce la rezultate distorsionate datorită şanselor mai mari de multicoliniaritate şi a valorilor erorilor standard mai ridicate (Jaccard, Turrisi, Wan, 1990).

Cea de-a doua metodă, denumită centrată, constă în înmulțirea celor doi termeni după ce în prealabil s-a scăzut fiecare valoare din media acestora. Tabelul de mai jos indică diferențe între cele două metode cu privire la modalitatea de obținere a variabilei interacțiune.

Tabelul 6.6. Exemplu de calculare a variabilei interacțiune prin metoda clasică (coloana a III-a), respectiv prin metoda centrată (coloana a VI-a)

X	Y	**X*Y**	X - X mediu	Y – Y mediu	**(X – X mediu)(Y – Y mediu)**
3	5	**15**	-1	0	**0**
2	6	**12**	-2	1	**-2**
5	3	**15**	1	-2	**-2**
6	7	**42**	2	2	**4**
1	5	**5**	-3	0	**0**
7	4	**28**	3	-1	**-3**

X mediu = 4; Y mediu = 5

Când se poate măsura acest efect de interacțiune?

În mod normal, efectul de interacțiune poate fi evidențiat, dacă ambele variabile sunt de tip numeric, măsurate fie prin scale de proporții, fie prin scale de interval. Unii autori recomandă evitarea măsurării efectului de interacțiune în cazul în care variabilele sunt de tip ordinal însă distorsiunile determinate de includerea unei asemenea variabile nu sunt foarte mari. Fără a intra în detalii tehnice, vom fi de acord cu poziția adoptată de Jaccard, Turrisi şi Wan (1990) care afirmă că includerea unor predictori măsurați prin scale ordinale în analiza de regresie este în sine un fapt controversat, iar analiza efectelor de interacțiune dintre variabilele în care cel puțin una este măsurată prin scală ordinală nu este cu nimic mai controversată decât analiza efectelor simple. Prin urmare, dacă cercetătorul va decide să-şi asume riscul obținerii unor rezultate distorsionate, prin includerea unor variabile măsurate prin scale de tip ordinal, va putea analiza şi efectul de interacțiune dintre acestea, asumându-şi aceleaşi riscuri.

În schimb, variabilele de tip categorial transformate în variabile de tip dummy pot fi supuse unei analize a efectelor de interacțiune, în condițiile în care una dintre variabile este măsurată prin scale de tip interval sau proporții. Dacă ambele variabile interacționate sunt de tip categorial se preferă însă o altă abordare, prin apelul la metoda clasică a tehnicilor ANOVA.

Un exemplu de măsurare a efectului de interacțiune dintre o variabilă numerică şi o variabilă de tip dummy poate fi sesizată în exemplul următor:

> Un cercetător doreşte să observe în ce măsură relația dintre volumul de muncă (gradul de solicitare) al angajaților şi stresul profesional perceput de aceştia este moderată de participarea acestora în cadrul unor programe de consiliere psihologică. Să presupunem că există trei loturi de subiecți: o parte care au urmat şedințe de consiliere cu psihologul companiei, o parte care a primit instrucțiuni pentru efectuarea unor exerciții de relaxare acasă şi un grup de control. Dacă nu există un efect de interacțiune între tipul de intervenție psihologică urmată şi gradul de solicitare la locul de muncă asupra stresului perceput, ar însemna că intervenția psihologică nu are un rol moderator în relația dintre cele două.

Relația simplă dintre gradul de solicitare la locul de muncă şi stresul perceput poate fi scrisă:

Stres = b0 + b1*Grad de solicitare

Aceasta poate fi calculată pentru fiecare tip de intervenție psihologică:

Stres 1 = b0 + b1*Grad de solicitare (grupul de subiecți supuşi consilierii psihologice)
Stres 2 = b0 + b1*Grad de solicitare (grupul de subiecți ce a urmat acasă exerciții de relaxare)
Stres 3 = b0 + b1*Grad de solicitare (grupul de control)

Dacă valorile pantei de regresie, adică ale lui b1, sunt asemănătoare înseamnă că nu există un efect de interacțiune determinat de tipul de intervenție psihologică propusă. Totuşi, o asemenea abordare descriptivă are un grad ridicat de subiectivitate, nefiind propice analizei de regresie (ex. nu ne putem da seama dacă b1 de 1,53, față de un b1 de 1,82 este diferit sau nu). În consecință, pentru a măsura efectul de interacțiune, ecuația de mai sus nu este adecvată.

Nici măcar o ecuație care să măsoare efectul simplu al variabilei tip de intervenție psihologică asupra stresului nu este propice pentru a vedea dacă există un efect de interacțiune:

Stres = b0 + b1*Grad de solicitare + b2*Dummy Consiliere + b3*Dummy relaxare

O asemenea ecuație de regresie, în care valoarea de referință pentru variabila dummy este constituită de grupul de control, ne oferă posibilitatea de a răspunde la întrebările dacă gradul de solicitare contribuie semnificativ la nivelul de stres perceput la locul de muncă, respectiv dacă tipul de intervenție psihologică contribuie semnificativ la nivelul stresului perceput la locul de muncă. Ea nu ne permite însă să vedem dacă efectul tipului de intervenție psihologică asupra nivelului de stres perceput este acelaşi, atât la oameni care au un grad de solicitare profesională ridicat, cât şi la oameni care au un nivel de solicitare profesională scăzut.

Prin urmare, pe lângă variabilele măsurate direct, se creează alte două variabile pentru a sesiza efectul de interacțiune între variabilele dummy şi variabila ce viza gradul de solicitare profesională.

Stres = b0 + b1*Grad de solicitare + b2*Dummy Consiliere + b3*Dummy Relaxare + b4*Grad de solicitare* Dummy Consiliere + b5*Grad de solicitare* Dummy Relaxare

Dacă există o singură variabilă dummy, ca în cazul variabilei „sexul persoanei", efectul de interacțiune implică includerea unei singure variabile

suplimentare, față de modelul clasic, al efectelor simple, la fel ca în ecuațiile de mai jos:

Stres = b0 + b1*Grad de solicitare + b2*Dummy Sex (1)
Stres = b0 + b1*Grad de solicitare + b2*Dummy Sex + b3*Grad de solicitare*Dummy Sex (2)

Cum se poate cunoaşte dacă există un efect de interacțiune?

Metoda obişnuită de verificare a prezenței sau absenței acestui efect este regresia ierarhică, prin compararea modelelor estimative care includ sau nu efectul de interacțiune. De exemplu, dacă ecuația notată cu unu ar avea un $R^2 = .23$, iar cea notat• cu doi ar avea un $R^2 = .31$, atunci diferența de .08, adică 8% din dispersia variabilei criteriu, este pusă pe seama efectului de interacțiune. Această creştere ce reprezintă mărimea efectului variabilei interacțiune este semnificativă statistic dacă F schimbare (în engleză „F change") asociat acestei diferențe explicative între cele două modele este semnificativ statistic. Cu alte cuvinte, dacă F schimbare de la modelul 1 la modelul 2 ar avea o valoare de genul $F(1, 134) = 8.29$, $p < .01$, am putea afirma că există un efect de interacțiune între cei doi predictori, care aduce un plus explicativ evoluției criteriului.

Cum se interpretează coeficienții corespunzători unui efect de interacțiune?

Interpretarea efectelor simple

În cazul regresiei clasice, în care nu erau incluse efectele de interacțiune, interpretarea coeficienților b sau β se făcea simplu. De exemplu, coeficientul b corespunzător gradului de solicitare în muncă indica nivelul de modificare al criteriului (al stresului perceput) atunci când gradul de solicitare creşte cu o unitate în condițiile eliminării influenței variabilei „sexul persoanei". Cu alte cuvinte, coeficientul arăta cum se modifică stresul atunci când nivelul de solicitare în muncă creşte cu o unitate, păstrând la valori constante ceilalți predictori. Similar, coeficientul β indica nivelul de modificare al stresului, exprimat în abateri standard, atunci când nivelul de solicitare în muncă creşte cu o abatere standard, în condițiile păstrării constante a celorlalți predictori.

De asemenea, b corespunzător sexului persoanei indică valoarea cu care se modifică stresul perceput atunci când variabila se modifică cu o unitate, în condițiile în care nivelul de solicitare în muncă ia valori constante. Cum este vorba de o variabilă dummy (0 – femei; 1 – bărbați), coeficientul b indică diferența în ceea ce priveşte stresul perceput între femei şi bărbați, în condițiile păstrării constante a influenței nivelului de solicitare în muncă asupra stresului. Aşadar, se consideră că modificarea reprezentată prin valorile b sau β este aceeaşi indiferent de nivelul de solicitare în muncă.

În ecuațiile de regresie care includ un factor de interacțiune, valorile b şi β corespunzătoare efectelor simple îşi modifică puțin modul de interpretare. Astfel, coeficientul de regresie specific nivelului de solicitare în muncă este văzut ca

intensitatea cu care se modifică criteriul atunci când nivelul de solicitare creşte cu o unitate, în condițiile în care ceilalți predictori iau valoarea zero. Concretizând, vom spune că b asociat nivelului de solicitare în muncă ne arată cu cât se modifică stresul perceput atunci când nivelul de solicitare creşte cu o unitate, în condițiile în care sexul persoanei ia valoarea zero, adică în cazul subiecților de sex feminin. În cazul variabilei sexul persoanei, interpretarea devine: b indică diferența dintre stresul perceput de femei şi bărbați atunci când nivelul de solicitare în muncă ia valoarea zero.

Se impun două precizări. În cazul interpretării efectelor simple în regresii care includ efecte de interacțiune, b nu indică valoarea cu care se modifică criteriul atunci când predictorul respectiv creşte cu o unitate, indiferent de nivelul celorlalți predictori, ci arată modul în care se modifică criteriul atunci când predictorul asociat creşte cu o unitate, în condițiile în care ceilalți predictori iau valoarea zero. Deosebirea apare, deoarece valorile b nu sunt constante pe toată amplitudinea predictorilor cu rol moderator, ea fiind calculată doar în situația în care aceştia iau valoarea zero. Uneori valoarea are o semnificație reală, ca în cazul variabilelor dummy, în care cota 0 indică prezența subiecților de sex feminim, în alte situații însă ea nu are o semnificație reală, precum se întâmplă în cazul nivelului de solicitare în muncă, ale cărui valori sunt cuprinse între 1 şi 15.

Cea de-a doua precizare se referă la interpretarea acestei valori zero, în funcție de tipul de metodă adoptată în obținerea variabilei interacțiune. Dacă se foloseşte metoda clasică, bazată pe produsul valorilor variabilelor de interacțiune, zero nu are o semnificație specială însă dacă se foloseşte metoda centrată, valoarea zero reprezintă, de fapt, media variabilelor interacțiune. Într-o asemenea situație, interpretarea inițială a lui b (nivelul cu care se modifică criteriul atunci când predictorul asociat se modifică cu o unitate în condițiile în care ceilalți predictori iau valoarea zero) devine nivelul cu care se modifică criteriul atunci când predictorul asociat se modifică cu o unitate, iar ceilalți predictori iau valori medii. Din această perspectivă, interpretarea îşi pierde sensul avut în cazul variabilei sexul persoanei (în care valoarea medie nu semnifică nimic) însă câştigă în relevanță în cazul variabilei nivelul de solicitare în muncă, fiind vorba de o interpretare făcută la un nivel mediu de solicitare în muncă. De aceea, alegerea modului de construire a variabilei interacțiune poate ține seama şi de uşurința interpretării rezultatelor în funcție de metoda aleasă.

Interpretarea efectelor de interacțiune

Vom utiliza ca exemplu aceeaşi ecuație, cu următoarele valori asociate:

Stres = b0 + b1*Grad de solicitare + b2*Dummy Sex + b3*Grad de solicitare*Dummy Sex
Stres = 4,5 + 0,82*Grad de solicitare + 0,12*Dummy Sex + 0,18*Grad de solicitare*Dummy Sex

Interpretarea coeficientului de regresie b3, corespunzător efectului de interacțiune, indică nivelul în care se modifică b1 atunci când variabila sex se

modifică cu o unitate. Tehnic, am putea spune că acest coeficient indică şi nivelul cu care se modifică b2 atunci când variabila grad de solicitare se modifică cu o unitate. De obicei, în interpretare se ține seama de aceea dintre variabile care joacă un rol moderator. Cum în situația de față variabila moderatoare este sexul persoanei, ne vom axa pe prima posibilitate de interpretare. Astfel, vom putea afirma că stresul perceput se modifică în funcție de gradul de solicitare în muncă în mai mare măsură în cazul bărbaților decât în cazul femeilor, deoarece în cazul bărbaților b1 devine 1 (0,82+0,18), în timp ce în cazul femeilor el este de 0,82. În plan psihologic, am putea afirma că nivelul de stres perceput este în mai mare măsură dependent de nivelul de solicitare în muncă în cazul bărbaților, decât în cazul femeilor. Ca şi în alte situații, o asemenea interpretare are sens doar dacă coeficientul de regresie asociat efectului de interacțiune este semnificativ statistic.

6.7.3. Analiza de cale

Deşi regresia, în multiplele ei forme, se dovedeşte o tehnică statistică robustă, ce poate fi utilizată într-o multitudine de situații, realitatea psihologică complexă este uneori inadecvată pentru a permite utilizarea metodelor de regresie, deşi avem de-a face cu un design corelațional.

De exemplu, dacă vom dori să testăm teoria comportamentului planificat, ca extensie a teoriei acțiunii raționale elaborate de Fishbein şi Ajzen, vom constata că regresia nu poate fi utilizată decât printr-un artificiu tehnic (Hankins, French şi Horne, 2000). Conform teoriei planificării comportamentului, prezentată în figura 6.16, o persoană adoptă un anumit comportament, dacă are intenția de a-l adopta şi dacă consideră că are capacitatea de a derula acel comportament. De asemenea, intenția de a acționa într-un anumit fel este dependentă de trei factori: atitudinea subiectului față de problema respectivă, normele existente care aprobă sau dezaprobă un anumit comportament şi părerea subiectului cu privire la capacitatea sa de a adopta un anumit comportament. Conform acestei teorii, dacă un student decide să şoptească unui coleg de-al său răspunsurile la examen face acest lucru, pentru că are intenția de a-l ajuta şi pentru că este de părere că poate face acest lucru fără probleme (are controlul acțiunii). De asemenea, intenția sa de a-l ajuta se bazează pe atitudinea pe care o are față de acesta (ex. „este o persoană de treabă"), pe normele subiective existente în grup (ex. „este bine să-ți ajuți colegii", „este bine să fii solidar cu aceştia") şi pe credința că poate face acest lucru fără a fi observat şi sancționat.

Dacă ar fi să testăm validitatea acestui model prin intermediul regresiei, ar trebui să împărțim modelul în două secțiuni, fiecare fiind supusă apoi unei analize de regresie. Prima secțiune ar avea drept variabilă dependentă intenția de a acționa, iar ca predictori: normele subiective, atitudinea şi controlul acțiunii. Cea de

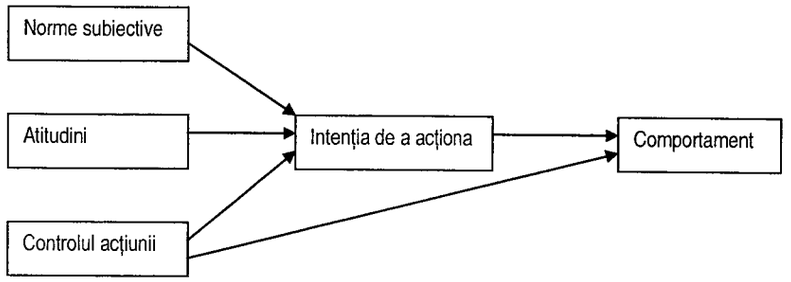


Figura 6.16. *Schema cauzală a teoriei comportamentului planificat*

a doua secțiune ar avea drept variabilă dependentă comportamentul adoptat, iar ca variabile independente: intenția de a acționa şi controlul acțiunii. Cele două secțiuni sunt sesizate clar în figura 6.17.

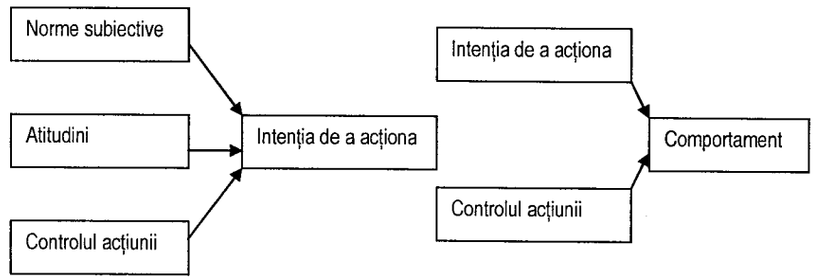


Figura 6.17. *Secțiuni ale teoriei comportamentului planificat testabile prin intermediul regresiei*

În termenii ecuației de regresie sub formă standardizată, cele două secțiuni pot fi scrise sub forma:

Intenția de a acționa = βNorme subiective + βAtitudini + βControlul acțiunii
Comportament = βIntenția de a acționa + βControlul acțiunii

Din păcate, chiar dacă prin regresie putem observa în ce măsură cele două ecuații de regresie explică într-un mod eficient cele două părți ale teoriei comportamentului planificat, nici una dintre ele nu oferă informațiile necesare cu privire la validitatea de ansamblu a teoriei propuse. Se pot face doar presupuneri, pornind de la ideea că, dacă ambele secțiuni sunt sprijinite de rezultatele analizei de regresie, atunci întregul model este acceptat.

În asemenea situații, în care modelul testat presupune mai multe variabile dependente, am putea utiliza **analiza de cale**, ca modalitate mai eficientă de a testa modelele cauzale complexe.

Definirea variabilelor în analiza de cale

Această metodă statistică este utilă în testarea unor modele cauzale, precum este cel prezentat în figura 6.16, vizând teoria comportamentului planificat. Într-un asemenea model există două tipuri principale de variabile: variabile **exogene** şi variabile **endogene** (Klem, 1995). De precizat că unii factori pot juca ambele roluri. Variabilele exclusiv exogene sunt variabilele de la care se porneşte schema cauzală, fiind similare predictorilor din regresia multiliniară. În figura 6.16, normele subiective, atitudinea şi controlul comportamentului sunt variabile exclusiv exogene, în timp ce intenția de a acționa joacă rolul de variabilă exogenă în relația pe care o are cu comportamentul adoptat. În schimb, aceeaşi variabilă, intenția de a acționa, joacă rolul de variabilă endogenă (similară criteriului din regresia multiliniară), deoarece ea este considerată ca fiind cauzată de acțiunea comună a trei variabile: normele subiective, atitudinea şi controlul comportamentului. De asemenea, comportamentul adoptat este o variabilă exclusiv endogenă, deoarece ea nu influențează alte variabile însă este influențată de intenția de acțiune şi de controlul comportamentului.

Printr-un asemenea model putem stabili în detaliu relația dintre variabile. Spre exemplu, putem stabili că există o relație indirectă între atitudinea unei persoane şi comportamentul acesteia, relația fiind mediată de intenția de a acționa. Similar, putem afirma că percepția subiecților despre uşurința sau dificultatea de a acționa (controlul comportamentului) are atât o acțiune directă asupra comportamentului adoptat, cât şi una indirectă, prin intermediul intențiilor de a acționa. De asemenea, putem compara două sau mai multe modele concurente pentru a vedea care dintre ele explică mai bine datele culese. Spre exemplu, prin analiza de cale se poate estima dacă teoria comportamentului planificat, prezentată în figura 6.16 explică mai bine comportamentul adoptat de o persoană în comparație cu o variantă modificată a acesteia (figura 6.18) sau față de teoria acțiunii raționale (figura 6.19).

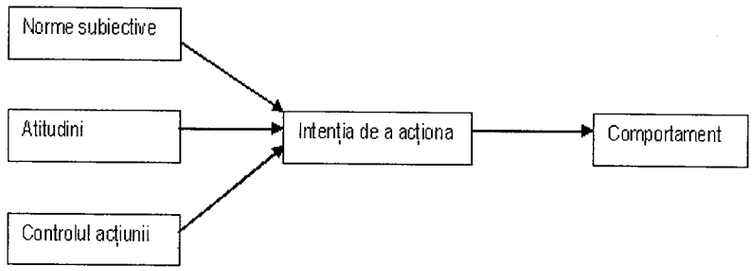


Figura 6.18. *Teoria comportamentului planificat în variantă modificată*

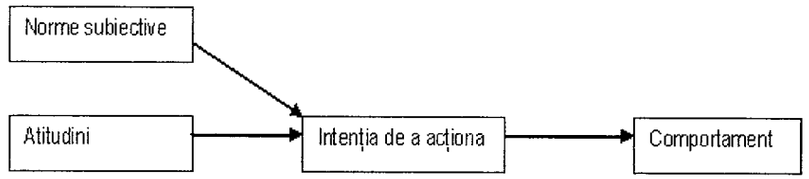


Figura 6.19. *Teoria acțiunii raționale*

Fără a insista mai mult asupra modului de realizare şi interpretare a analizei de cale şi a coeficienților rezultați, vom menționa că în selectarea modelului care explică cel mai bine datele obținute se foloseşte aceeaşi strategie şi aceiaşi parametri, utilizați şi în cazul analizei factoriale confirmatorii, interpretarea făcându-se într-un mod similar (vezi secțiunea 5.2.3.). Din acest punct de vedere, SPSS-ul nu oferă posibilitatea de a realiza direct o analiză de cale însă un program complementar, AMOS, poate realiza asemenea analize.

BIBLIOGRAFIE

Berry, W.D., Feldman, S. (1985). *Multiple regression in practice.* (Sage University Papers on Quantitative Applications in the Social Sciences, series no. 07-050). Newbury Park, CA: Sage.

Field, A. (2000). *Discovering statistic using SPSS for Windows.* London: Sage Publications.

Garson (2000). Găsit la 14.IX pe www2.chass.ncsu.edu/garson/pa765/regression.htm

Hankins, M., French, D., Horne, R. (2000). Statistical guidelines for studies of the theory of reasoned action and the theory of planned behaviour. *Psychology and Health*, 15, 151-161.

Jaccard, J., Turrisi, R., Wan, C.K. (1990). *Interaction effects in multiple regression.* (Sage University Papers on Quantitative Applications in the Social Sciences, series no. 07-072). Newbury Park, CA: Sage.

Kirby, J.R., Parrilla, R.K., Pfeiffer, S.L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, 95(3), 453-464.

Klem, L. (1995). *Path Analysis.* În L.G. Grimm şi P.R. Yarnold (eds.). Reading and understanding multivariate statistics. (p. 65-98). Washington, DC: American Psychological Association.

Licht, M.H. (1995). *Multiple regression and correlation.* În L.G. Grimm şi P.R. Yarnold (eds.). Reading and understanding multivariate statistics. (p. 19-64). Washington, DC: American Psychological Association.

Miles, D.E., Johnson, G.L. (2003). Aggressive driving behaviors: are there psychological and attitudinal predictors? *Transportation Research Part F*, 6, 147-161.

Nicol, A.A.M., Pexman, P.M. (1999). *Presenting your findings. A practical guide for creating tables*. Washington DC, USA: American Psychological Association.

Pedhazur, E.J., Schmelkin, L. (1991). *Measurement, design, and analysis: an integrated approach*. Hillsdale, USA: Lawrence Erlbaum Associates, Inc.

Pitariu, H. (1983). *Psihologia selecției și formării profesionale*. Cluj-Napoca: Editura Dacia.

Schroeder, L.D., Sjoquist, D.L., Stephan, P.E. (1986). *Understanding regression analysis. An introductory guide*. (Sage University Papers on Quantitative Applications in the Social Sciences, series no. 07-057). Newbury Park, CA: Sage.

Thompson, B. (1995a). Stepwise regression and stepwise discriminant analysis need not apply here: a guidelines editorial. *Educational and Psychological Measurement*, 55(4), 525-534.

Thompson, B. (1995b). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. *Educational and Psychological Measurement*, 55(1), 84-94.

CAPITOLUL 7

REGRESIA LOGISTICĂ

Dacă criteriul sau variabila dependentă inclusă într-o regresie nu este măsurată printr-o scală numerică, ci printr-una dihotomică de tipul „da" sau „nu"; „admis" sau „respins"; „corect" sau „greșit", „succes" sau „insucces școlar" etc., se aplică o altă formă de regresie, denumită regresie logistică.

Similar regresiei liniare, regresia logistică poate fi simplă sau multiplă, cu scop predictiv sau cu scop explicativ, cuprinzând doar efecte simple sau și pe cele de interacțiune. Modul de lucru în asemenea cazuri este analog cu cel prezentat în capitolul destinat regresiei liniare. De exemplu, și în cazul regresiei logistice putem întâlni atât variabile numerice, cât și variabile dummy, ecuații de regresie liniare sau neliniare, modele simultane, ierarhice sau în pași etc. De aceea nu vom mai prezenta toate aceste detalii, ci ne vom concentra asupra modului de realizare și asupra particularităților regresiei logistice.

Un exemplu tipic în care se poate utiliza regresia logistică este cazul infidelității în plan sexual. În funcție de acest criteriu am putea împărți oamenii în fideli sau infideli față de parteneri. Dacă ar fi să găsim un set de variabile care să realizeze o predicție cât mai corectă a comportamentului în cuplu al partenerilor, am putea utiliza regresia logistică în scop predictiv. În schimb, dacă am fi interesați să vedem ce factori psihologici pot explica infidelitatea după ce am controla unii factori personali sau sociali precum vârsta, sexul subiectului, statutul subiectului, venitul realizat, nivelul de educație, oportunitățile existente la locul de muncă etc., am utiliza regresia logistică în scop explicativ.

7.1. Concepte și principii teoretice

Așadar, regresia logistică este o formă de regresie în care variabila dependentă (VD) este dihotomică, în timp ce variabilele independente (VI) pot fi continue sau categoriale. Există cazuri în care VD, deși este categorială, are mai multe alternative. Un exemplu ar fi identificarea unor predictori care să explice cât mai mult din evoluția probabilității ca oamenii să voteze pe unul dintre liderii politici precum Adrian Năstase, Teodor Stolojan sau Corneliu Vadim Tudor. În această situație vorbim despre o regresie logistică multinomială. În capitolul de față vom aborda doar regresia logistică dihotomică sau binară, denumită generic „regresie logistică". Modalitatea de interpretare a regresiei logistice multinomiale, cât și a regresiei logistice ordinale nu poate fi înțeleasă fără o bună aprofundare a regresiei

logistice binare. Mai multe detalii despre aceste modalități de regresie logistică pot fi găsite în Agresti (1990).

Asemeni regresiei liniare, cea logistică încearcă să găsească un model explicativ cât mai bun, exprimat printr-o funcție matematică. Regresia logistică se aseamănă cu analiza logit, cele două tehnici oferind chiar rezultate identice, însă analiza logit are un caracter mai restrictiv, deoarece toți predictorii trebuie să fie măsurați prin scale de tip nominal. De asemenea, regresia logistică poate fi folosită cu succes în locul analizei discriminantului, care este o tehnică mult mai pretențioasă din perspectiva condițiilor necesare pentru a fi utilizată corect. Se cuvine să precizăm şi câteva diferențe existente între regresia liniară şi regresia logistică.

(a) Spre deosebire de regresia liniară, nu există condiția de liniaritate între VI şi VD, nu există cerința ca VI să fie distribuite normal, fiind astfel mai uşor de aplicat. Din toate aceste puncte de vedere regresia logistică se dovedeşte a fi o metodă foarte robustă şi utilă.

(b) Din punct de vedere tehnic regresia logistică se bazează pe metoda estimării verosimilității maxime (în engleză „maximum likelihood estimation") şi nu pe metoda clasică, a celor mai mici pătrate.

(c) În fine, dacă în regresia liniară cunoaşterea VI duce la predicția valorii VD, în regresia logistică cunoaşterea VI duce la determinarea (predicția) unei probabilități pentru fiecare dintre cele două răspunsuri posibile ale VD. În cazul regresiei liniare se obține o valoare estimată a criteriului, pentru fiecare subiect în parte, pe baza ecuației de regresie rezultate. De exemplu, am putea estima performanța în vânzări a unei persoane pe baza experienței profesionale a acesteia şi pe baza abilităților sale psihologice. În cazul regresiei logistice, dat fiind faptul că nu există decât două răspunsuri posibile (ex. fidel sau infidel față de partener), se poate estima probabilitatea, pentru fiecare subiect în parte, de a se comporta într-un mod fidel sau infidel, iar acest lucru ar fi obținut pe baza ecuației de regresie logistică. Deşi are aparent o formă liniară, criteriul estimat este o variabilă logit, denumire ce vine de la faptul că VD din ecuație este logaritmul natural al unei valori ce intră în calculul probabilității de a se întâmpla un eveniment sau altul. În cazul infidelității, criteriul indică probabilitatea ca o persoană să se comporte într-un mod fidel în plan sexual sau dimpotrivă.

Precizări metodologice necesare aplicării regresiei logistice

Categoriile VD trebuie să fie reciproc exclusive. Fidel şi infidel în plan sexual față de partener pot constitui categorii ale VD în regresia logistică, deoarece ele se exclud reciproc. În schimb, fan al fotbalului şi al tenisului sunt categorii care nu se exclud reciproc, deoarece există persoane care pot intra în ambele categorii sau dimpotrivă, în nici una. Prin urmare, regresia logistică nu poate fi aplicată într-o asemenea situație.

Datele trebuie să fie independente unele de altele. Dacă avem categoriile bolnav, respectiv sănătos psihic, nu putem include datele de la acelaşi subiect

obținute la intervale de timp diferite, în perioada episodului psihopatologic, respectiv după însănătoşire.

Modelul logistic trebuie să cuprindă *VI semnificative din punct de vedere teoretic pentru relația cu VD.* Includerea unor VI care nu au legătură cu VD poate conduce la erori de estimare a coeficienților logit. De asemenea, pot apărea corelații false între VI şi VD, caz în care este afectată calitatea modelului.

Utilizarea metodei de estimare a verosimilității maxime necesită *un număr mai mare de subiecți* decât utilizarea metodei celor mai mici pătrate. Din acest motiv, pentru eşantioane mici, testarea ipotezelor statistice poate fi inadecvată. Un criteriu conservator susține necesitatea unui minimum de 50 de subiecți pentru fiecare variabilă predictor. Totuşi, pentru cercetări de peste 300 de subiecți, acest raport poate să scadă, situându-se undeva în jurul a 20-30 de subiecți pentru fiecare predictor (Wright, 1995).

7.2. Modelul regresiei logistice

Pentru a clarifica principiile teoretice ale regresiei logistice vom apela în paralel la un exemplu din psihologia politică. Mai exact, vom surprinde în ce măsură vârsta influențează opiniile românilor în ceea ce priveşte întrebarea: "Era mai bine înainte de 1989 decât după 1989, din perspectiva conducătorilor țării?". Datele provin din Barometrul de Opinie Publică din octombrie 1999, răspunsurile subiecților fiind modificate prin a reduce alternativele de răspuns de la cinci la două categorii. Aceste variante sunt: "1 – era mai bine înainte de 1989"; "0 – era mai rău înainte de 1989".

Dacă modelul ecuației de regresie liniare simple este:

$$Y = b0 + b1*X \qquad (7.1)$$

modelul regresiei logistice simple este (Menard, 1995):

$$\ln(y) = b0 + b1*X \qquad (7.2)$$

care poate fi scris şi

$$y = \exp(b0 + b1*X) \qquad (7.3)$$

unde y este şansa sau raportul dintre probabilitatea de a răspunde 1 (era mai bine înainte de 1989) şi probabilitatea de a răspunde 0 (era mai rău înainte de 1989). Adică, ln (y) = ln [p/(1-p)]; b0 şi b1 sunt coeficienții logit corespunzători coeficienților de regresie (b0 fiind asociat interceptului, iar b1 fiind asociat pantei de regresie determinate de predictorul X, în acest caz, vârsta subiecților).

Raportul ia valoarea 1 când există o egalitate între cele două probabilități, adică atunci când fiecare răspuns are o probabilitate de .50 (50%) de a fi ales de o anumită persoană. El devine supraunitar când există mai multe şanse să se întâmple situația 1 (subiectul să afirme că era mai bine înainte de 1989) şi subunitar când şansele sunt mai mari de a prezice varianta 0 (era mai rău înainte de 1989).

Devine necesară diferențierea clară dintre *probabilități, şanse şi raportul şanselor.* În regresia logistică dihotomică, *şansele* unui subiect *de a intra în categoria 1* (era mai bine înainte de 1989) sunt *egale cu raportul dintre probabilitatea de a fi inclus în acea categorie şi probabilitatea de a aparține celeilalte categorii* (0 - era mai rău înainte de 1989). Raportul se inversează dacă dorim să calculăm şansele cazului de a intra în categoria 0. Dacă probabilitatea de a fi inclus în categoria 1 este p = .80, ea devine .20 pentru categoria 0 (q=1-p). Şansele ca subiectul să aparțină primei categorii sunt de patru ori mai mari (.80/.20). Se observă că *probabilitățile pot lua valori între 0 şi 1*, în timp ce *şansele variază de la 0 la infinit.* Prin urmare, se pune problema calculării probabilității ca un anumit subiect să fie inclus în prima categorie. Ea se face după formula (Wright, 1995):

$$p = \frac{e^{y}}{1 + e^{y}} \qquad (7.4)$$

unde p = probabilitatea de a se petrece evenimentul 1 ("era mai bine înainte de 1989"); e simbolizează exponentul, iar y ia valori diferite în funcție de tipul de ecuație logistică. Astfel:

y = b0 + b1*X pentru regresii logistice simple
y = b0 + b1*X1 + b2*X2 ... pentru regresii logistice multiple

Vom calcula, ilustrativ, probabilitatea unei persoane de 20 de ani să afirme că era mai bine înainte de 1989. Să presupunem că interceptul (b0) a fost –1,02, iar coeficientul logit al variabilei vârsta (b1) a fost 0,03. Vom obține: y = -1,02 + 0,03*20 = -0,42

Deoarece y = exp (b0 + b1*X), exponentul valorii y de -0,42 este 0,657. Putem înlocui în formulă pentru a afla probabilitatea ca un tânăr de 20 de ani să afirme că era mai bine înainte de 1989.

$$p = \frac{0,657}{1+0,657} = 0,396$$

Valoarea ne determină să credem că tinerii de 20 de ani vor afirma mai degrabă că este mai bine acum comparativ cu situația dinainte de 1989, deoarece probabilitatea de a afirma că era mai bine înainte de 1989 din perspectiva

conducătorilor țării era de 0,396, în timp ce probabilitatea de a se opune unui asemenea răspuns era 1-p, adică 0,604, o valoare mai mare.

Repetând acelaşi demers matematic pentru o persoană de 50 de ani vom obține:

y = -1,02 + 0,03*50 = 0,48, iar exp (y) = 1,61
p = 1,61 / (1+1,61) = 0,616

În acest caz, pentru o persoană în vârstă de 50 de ani există o probabilitate mai mare de a afirma că era mai bine înainte de 1989 (p = .616) decât de a susține contrariul (q = 1-p = .384).

Decizia de a fi inclus într-o categorie sau alta se face având criteriu limită o probabilitate de .50 însă această valoare poate fi modificată în funcție de caracteristicile VD şi de frecvența reală a alternativelor de răspuns (Hair, Anderson, Tatham şi Black, 1998).

Pe baza probabilităților putem stabili şansele de a fi inclus într-o categorie sau alta. Un tânăr de 20 de ani va avea de aproximativ 1,5 ori mai multe şanse de afirma că este mai bine acum decât înainte de 1989 (.604/.396). În schimb, o persoană de 50 de ani va avea tot de aproximativ 1,5 ori mai multe şanse să afirme că era mai bine înainte de 1989 decât acum (.616/.384).

Un concept esențial în regresia logistică este cel de *raport al şanselor.* El *estimează schimbările intervenite în şansa de a fi inclus în prima categorie* (în cazul dat: era mai bine înainte de 1989), *atunci când variabila predictor creşte cu o unitate.* Pentru a-l calcula avem nevoie doar de coeficientul logit al variabilei predictor în cauză.

Să presupunem că am obținut un parametru b = 0,50 pentru un alt predictor: numărul de copii. Pentru a afla raportul şanselor, vom calcula exponent din valoarea acestuia (0,50).

$$e^{0,50} = 1,64$$

Interpretarea raportului de şanse va fi prezentată pe larg când vom discuta despre coeficienții logit. Până atunci ne vom rezuma să citim rezultatul prezentat mai sus, afirmând că: un individ cu doi copii are de 1,64 ori mai multe şanse să afirme că era mai bine înainte de 1989 în comparație cu o persoană cu un singur copil (VI creşte cu o unitate). În cazul mai multor VI incluse în model se specifică sintagma "păstrând celelalte variabile constante".

Se poate calcula raportul şanselor şi pentru modificări ale VI mai mari de o unitate. Astfel, vom compara o persoană cu cinci copii cu o alta care are un copil. Diferența de patru unități va fi exprimată astfel:

$$e^{4*0,50} = 7,38$$

Adică, un subiect cu 5 copii, în comparație cu cel care are un singur copil, va avea de 7,38 ori mai multe şanse de a afirma că era mai bine înainte de 1989, din perspectiva conducătorilor țării.

În privința rezultatelor oferite de SPSS, lucrurile sunt diferite. SPSS-ul oferă direct valoarea raportului şanselor pentru fiecare VI atunci când acestea se modifică cu o unitate şi dă posibilitatea de a opta pentru afişarea intervalului de încredere corespunzător fiecărui raport al şanselor. În schimb, programul nu prezintă direct şansele estimate pentru fiecare subiect. Acestea pot fi deduse indirect pe baza probabilităților obținute. Deoarece probabilitățile se calculează pentru fiecare subiect în parte, există în SPSS opțiunea de a salva ca variabilă nouă, probabilitatea fiecărui subiect de a răspunde 1 (de a afirma că era mai bine înainte de 1989).

Vom sublinia că în *interpretarea coeficienților logit obținuți ținem seama de raportul de şanse şi nu de probabilități.* Acestea, nu sunt eficiente deoarece au valori diferite la variații constante ale variabilei predictor. În general, probabilitățile se modifică mai mult pentru valori medii ale VI şi mai puțin pentru valori extreme ale VI.

Astfel, dacă admitem un intercept b0 = -1,02 şi un coficient logit b1 = 0,50 pentru variabila număr copii, vom observa că probabilitățile nu se modifică uniform când VI creşte cu o unitate. Concret, probabilitatea unei persoane cu un copil de a afirma că era mai bine înainte de 1989 este .37, iar cea a unei persoane cu doi copii este .49, diferența de probabilitate dintre cele două răspunsuri fiind de .12. În schimb, probabilitatea unui subiect cu cinci copii devine .81 față de .87 cât este probabilitatea pentru o persoană cu şase copii, diferența fiind de numai .06.

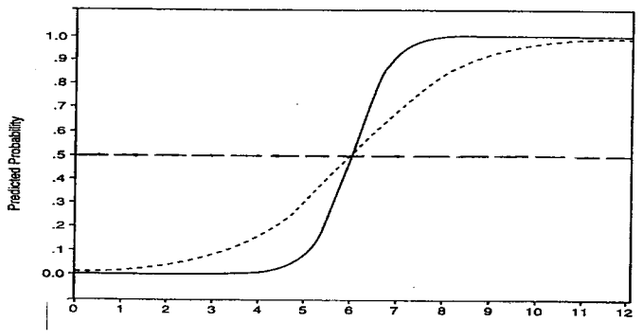


Figura 7.1. *Modalități de evoluție a curbei regresiei logistice simple în trei situații ipotetice (b = 3 – linie continuă; b = 1,5 – linie punctată; b = 0 – linie întreruptă)*

Acest lucru se întâmplă, deoarece regresia logistică nu are o formă liniară, ci o formă specifică în formă de S, denumită curba de regresie logistică. Fiecare combinație de coeficienți logit cu interceptul produce o modificare a formei acesteia. În funcție de valorile coeficientului logit al VI se modifică panta curbei şi

direcția acesteia. Figura de mai jos prezintă forma curbei logistice pentru trei valori diferite ale coeficientului logit pornind de la ideea că interceptul a luat valoarea 0.

Se poate observa că panta este mai abruptă în cazul liniei continue, deoarece valoarea coeficientului logit a fost mai îndepărtată de 0 (b=3). Panta scade în cazul liniei punctate, pentru că valoarea coeficientului logit este mai apropiată de 0 (b=1,5). Panta arată cât de bine poate discrimina coeficientul logit luat în calcul. Cu cât este mai abruptă, cu atât variabila predictor intervine mai mult în modificarea probabilităților VD. Un coeficient logit (b=0) nu va influența în nici un fel modificarea probabilităților atunci când variabila predictor va lua valori diferite. În această situație, curba se transformă într-o linie orizontală (linia întreruptă).

Pentru valori pozitive ale coeficientului logit, curba creşte dinspre stânga spre dreapta. Aceasta înseamnă că valorile mici ale predictorului determină o probabilitate mai mică de a se întâmpla evenimentul 1, iar valorile mari cresc această probabilitate. Situația se inversează dacă obținem un coeficient logit negativ. Curba de regresie va creşte dinspre dreapta spre stânga, ceea ce înseamnă că valorile mari ale predictorului determină o probabilitate mică de apariție a evenimentului 1 şi invers. Astfel, un b= -1,5 va determina o curbă inversă față de curba punctată rezultată din coeficientul logit b de 1,5, prezentat în figura 7.1. Ea va păstra însă celelalte caracteristici şi va avea o pantă mai abruptă comparativ cu modelul 3 (b=0), dar mai lină decât modelul 1 (b=3). De remarcat că niciodată curba nu va depăşi probabilitatea 1 şi nu va cădea sub probabilitatea 0, chiar dacă vom utiliza valori extreme ale VI (Wright, 1995).

7.3. Estimarea modelelor logistice

În această secțiune vom discuta paşii necesari pentru aplicarea regresiei logistice. Vom introduce, în paralel cu explicațiile conceptuale, următoarea situație-exemplu:

O serie de factori influențează naşterea unor copii sănătoşi sau, dimpotrivă, a unor copii cu malformații genetice diverse. Observând un număr de 125 de naşteri, un cercetător a anchetat mamele cu privire la 4 factori: vârsta mamei, existența în familie (rude de gradul I şi II) a unor malformații genetice, consumul de alcool în perioada sarcinii şi stresul cotidian perceput.

Baza de date aferentă acestui exemplu este regresie logistica.sav, iar pentru analiza datelor în SPSS 11.0 se urmează calea: „Analyze"-„Regression"-„Binary Logistic".

Regresia logistică oferă posibilitatea de a:

- testa dacă un model este semnificativ statistic;
- compara două modele între ele;
- stabili eficiența modelului ales;
- interpreta coeficienții logit semnificativi din modelul ales;
- stabili profiluri ideale pe baza ecuației logistice obținute;

- analiza cazurile influente din modelul explicativ găsit.

7.3.1. Testarea modelelor explicative logistice

Vor fi prezentate trei metode ce testează dacă un model este sau nu semnificativ statistic. Este vorba de testul G^2 al rapoartelor de verosimilitate (în engleză „chi-square likelihood ratio test"); testul Hosmer şi Lemeshow de măsurare a potrivirii (în engleză „Hosmer and Lemeshow's Goodness of Fit Test"); şi criteriul bayesian ajustat de informare notat cu BIC' (în engleză „Adjusted Bayesian Information Criterion"). Primele două metode sunt incluse în outputul SPSS-ului, iar ultima poate fi uşor calculată pornind de la datele prezentate de acelaşi program statistic.

A. **Testul G^2 al rapoartelor de verosimilitate** este o variantă a testului χ^2 cu proprietăți care-l fac mai potrivit pentru a identifica modelele explicative optime. Pentru a înțelege modul său de funcționare vom introduce câteva concepte. Deşi, la prima vedere, acestea par complicate, trecerea lor în revistă este necesară pentru a înțelege rezultatele oferite de SPSS, concretizate în indicatori precum: χ^2 inițial, χ^2 al pasului, G^2 etc.

Un prim concept este *log-likelihood* (LL), un logaritmul calculat prin iterații utilizând metoda estimării verosimilității maxime. Valorile LL sunt negative, fiind zero doar dacă verosimilitatea este 1.

Din acest considerent se preferă utilizarea *abaterii*, notată (-2LL), care este egală cu dublul logaritmului de mai sus (log-likelihood) înmulțit cu minus unu pentru a lua valori pozitive. Abaterea este modalitatea de lucru a SPSS-ului, iar pe baza ei pot fi calculați mai mulți indicatori.

Hi pătrat inițial măsoară erorile asociate cu modelul explicativ propus în situația în care singurul parametru inclus este interceptul. Se notează cu –2LLnull sau –2 log likelihood initial.

Hi pătrat ca măsură a nepotrivirii se notează cu –2LLmodel sau –2 log likelihood şi reflectă erorile asociate cu modelul explicativ propus când includem atât interceptul, cât şi toate VI dorite de cercetător.

Testul rapoartelor de verosimilitate se notează cu G^2 şi este hi pătrat al modelului. Este principalul indicator care ne arată în ce măsură modelul explicativ propus este sau nu semnificativ statistic. Matematic se obține utilizând formula:

$$G^2 = (-2LLnull) - (-2LLmodel) \qquad (7.5)$$

G^2 testează ipoteza nulă conform căreia nici o VI nu este asociată cu logaritmul şanselor ca VD să ia valoarea 0 sau 1. Cu alte cuvinte, testul verifică dacă toți coeficienții logit ai VI, exceptând interceptul, sunt zero. Aşadar, testul G^2 măsoară modelul în ansamblu şi nu se referă la fiecare coeficient logit în parte. Aceştia sunt verificați prin testul Wald sau alte probe similare.

Testul G^2 urmează o distribuție χ^2, unde numărul gradelor de libertate este dat de numărul de VI incluse în model. Dacă G^2 este semnificativ statistic ($p < .05$), respingem ipoteza nulă şi putem concluziona că modelul propus este superior celui care cuprinde doar interceptul, deoarece există cel puțin un predictor, în model, care influențează semnificativ logaritmul şanselor ca VD să se modifice spre 0 sau spre 1. G^2 este analog testului F din regresia liniară multiplă (Garson, 2000).

Un alt criteriu de testare a modelului explicativ presupune o acțiune inversă pentru a surprinde diferența semnificativă dintre un model saturat şi modelul propus. Aceasta este metoda preferată de lucru în analiza logliniară ierarhică.

B. **Testul Hosmer şi Lemeshow ca măsură a potrivirii** testează ipoteza nulă conform căreia datele observate sunt generate de modelul explicativ propus. Dacă testul se dovedeşte a fi semnificativ statistic ($p < .05$) înseamnă că datele generate de modelul propus sunt diferite de datele observate. Ca urmare, dorim ca rezultatul să fie nesemnificativ statistic, pentru a nu respinge ipoteza nulă ce afirma că nu există diferențe între valorile observate şi cele prezise de model. Testul nu se referă la intensitatea relației, ci doar demonstrează că există o anumită influență a VI din model asupra VD. Cu cât eşantionul devine mai mare, cu atât şansele ca ipoteza nulă să fie respinsă cresc (Garson, 2000).

Ambele modele prezentate mai sus sunt derivări ale lui χ^2, în consecință sunt dependente de volumul eşantionului. Un număr mare de subiecți (ex. 1500 subiecți) creşte şansa de a respinge ipoteza nulă în comparație cu un eşantion mai mic (ex. 200 subiecți). Acest fapt determină testul Hosmer şi Lemeshow să devină mai conservator pe măsură ce volumul eşantionului creşte, în timp ce testul G^2 devine mai liberal. Pentru a elimina aceste efecte, vom prezenta Criteriul bayesian de informare în forma sa ajustată (BIC'), dezvoltat pe baza principiilor statisticii bayesiene.

C. **Criteriul Bayesian Ajustat de Informare** (BIC') poate fi utilizat atât pentru testarea unui model anume, cât mai ales pentru a compara două modele diferite. Bineînțeles că există şi un Criteriu Bayesian de Informare (BIC) neajustat, care este însă afectat, în mai mare măsură decât varianta ajustată, de numărul de variabile incluse în model şi de volumul eşantionului. Formula de calcul al lui BIC' porneşte de la testul G^2 şi este:

$$BIC' = - G^2 + df' \text{ model} * \ln N \qquad (7.6)$$

unde df' se referă la numărul de grade de libertate ajustate, adică la numărul de parametri regresori din model (excluzând interceptul), iar N este numărul de subiecți din studiu.

Totuşi, BIC' avantajează modelele ce dau dovadă de parcimonie în dauna unor modele complexe. Acelaşi rol îl îndeplineşte şi un alt indicator AIC (Akaike Information Criterion), care însă este mai puțin conservator în avantajarea modelelor simple (Wickens, 1998).

Dacă BIC' ia valori pozitive, modelul nu explică foarte bine datele obținute. El este chiar mai slab în explicarea variației VD decât modelul nul, ce include numai interceptul. Vom discuta în cadrul secțiunii următoare modul în care este utilizat BIC' pentru a compara două modele între ele, după un algoritm propus de Long (1997) şi Freese (2000).

În exemplul amintit anterior, vizând factorii ce intervin în apariția unor malformații genetice la 125 de nou-născuți, vom evalua trei factori: consumul de alcool în perioada sarcinii, existența unor rude apropiate cu malformații genetice şi vârsta mamei la naştere. Ecuația logistică este:

ln Handicap = b0 + b1*Alcool + b2*Factori genetici + b3*Vârsta mamei

Pentru testarea modelului vom apela atât la output-ul oferit de SPSS cât şi la anumite calcule suplimentare.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	54,501	3	,000
	Block	54,501	3	,000
	Model	54,501	3	,000

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12,966	7	,073

unde Chi-Square Model este testul G^2.

Figura 7.2. *Testarea modelului propus din perspectiva gradului de semnificație*
SPSS este o marcă înregistrată.

Vom scrie rezultatul sub forma $G^2(3) = 54.501$, $p < .001$. Gradele de libertate sunt date de numărul VI incluse în modelul testat. Pornind de la aceşti indicatori poate fi calculat şi BIC', procedeu care nu este oferit direct în output-ul SPSS-ului.

Vom avea:

BIC' = -54,501 + 3 * ln 125 = - 40,01

Acest rezultat poate fi interpretat ca un credit dat modelului explicativ propus. O valoare pozitivă a lui BIC' ar fi însemnat că avem un model explicativ mai slab decât modelul nul, ce ține seama doar de intercept.

SPSS-ul oferă rezultate pentru cea de-a treia metodă de testare a modelelor logistice, testul Hosmer şi Lemeshow ca măsură a potrivirii. Pentru aceasta trebuie selectată opțiunea corespunzătoare în fereastra "Options" a regresiei logistice.

Rezultatul obținut arată că ipoteza nulă nu poate fi respinsă ($p > .07$), ca urmare vom putea spune că datele rezultate, estimate pe baza modelului explicativ, nu sunt semnificativ diferite de valorile observate. Toate aceste rezultate sprijină ideea că cel puțin un predictor din cei trei (consumul de alcool, factorii genetici şi vârsta mamei) contribuie semnificativ la estimarea probabilității de a da naştere unui copil sănătos sau unuia cu malformații genetice.

7.3.2. Compararea modelelor logistice

Înainte de a trece în revistă principalele modalități de a compara modele, trebuie subliniat că există două tipuri de relații între acestea: de tip "cuib", respectiv modele diferite (Hair et al., 1998; Garson, 2000).

Două modele logistice se află într-o relație de tip cuib, dacă într-un model sunt incluse toate VI care se găsesc în celălalt. În caz contrar, modelele sunt socotite a fi diferite. Pentru exemplul dat, o relație de tip "cuib" este:

Ln y = b0 + b1*Vârsta mamei + b2*Consumul de alcool şi
Ln y = b0 + b1*Vârsta mamei + b2*Consumul de alcool + b3*Stresul perceput

Pentru aceeaşi situație, comparația între modele diferite ar fi:

Ln y = b0 +b1*Vârsta mamei + b2*Consumul de alcool + b3*Stresul perceput şi
Ln y = b0 +b1*Vârsta mamei + b2*Antecedentele genetice

Pe baza acestei diferențieri stabilim tehnica necesară pentru compararea modelelor. Vom descrie două procedee: χ^2 al pasului şi diferența dintre criteriile bayesiene ajustate de informare. Ambele metode pot fi utilizate pentru modele de tip "cuib", iar ultima poate fi utilizată şi pentru compararea modelelor explicative diferite.

Hi pătrat al pasului este metoda utilizată în SPSS. Tehnica surprinde mărimea modificărilor în ceea ce priveşte log likelihood prin introducerea unor VI noi, în condițiile în care VI anterioare sunt păstrate. Se verifică astfel, dacă anumite VI merită a fi introduse în model, dacă au un efect suficient de mare, încât eliminarea lor ar conduce la un model explicativ inferior. Tehnica este eficientă şi în identificarea variabilelor dummy semnificative statistic, care pot fi testate distorsionat prin proba Wald. Din punct de vedere tehnic, χ^2 al pasului este analog lui F schimbare din cadrul metodei ierarhice din regresia liniară multiplă.

Regresia logistică poate fi calculată şi prin alte metode decât cea simultană (în engleză „enter") ori cea ierarhică. SPSS-ul poate utiliza atât metoda

progresivă "stepwise", cât şi pe cea regresivă "backwise". În aceste situații este oferit un χ^2 "improvement", similar ca funcție lui χ^2 al pasului. Dacă modelele comparate nu se află într-o relație de tip cuib, nu vom putea folosi χ^2 al pasului oferit de SPSS, ci va trebui să apelăm la valorile BIC'.

Diferența dintre valorile BIC' a două modele poate fi utilizată indiferent dacă modelele se află sau nu într-o relație de tip cuib. Pentru aceasta trebuie întâi calculat BIC' pentru fiecare model, iar apoi făcută diferența dintre cele două valori: BIC'1 – BIC'2. Dacă BIC'1 – BIC'2 > 0, vom prefera modelul al doilea; dacă BIC'1 – BIC'2 < 0, vom prefera primul model.

Raftery, autorul conceptelor BIC şi BIC', propune o grilă empirică de interpretare a diferențelor ce este prezentată în tabelul 7.1 (apud Freese, 2000). De pildă, dacă diferența dintre două modele în ceea ce priveşte BIC' este 8, putem afirma că modelul cu BIC' mai mic este mult mai bun, existând o probabilitate de .95 - .99 ca acest fapt să fie adevărat.

Tabelul 7.1. Diferențele critice între modele după Raftery (apud, Freese, 2000)

Diferența în valoarea absolută	Semnificație	Probabilitatea implicată
0 – 2	Slabă	.50 - .75
2 – 6	Medie	.75 - .95
6 – 10	Puternică	.95 - .99
> 10	Foarte puternică	> .99

Vom utiliza aceeaşi problemă pentru exemplificare. Modelele comparate vor fi:

ln Handicap = b0 + b1*Alcool + b2*Factori genetici + b3*Vârsta mamei
ln Handicap = b0 + b1*Alcool + b2*Factori genetici + b3*Vârsta mamei + b4*Stresul perceput

Primul model a fost discutat şi testat anterior, valoarea obținută fiind $G^2(3) = 54,501$.

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	54,501	3	,000
	Block	54,501	3	,000
	Model	54,501	3	,000

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,034	1	,855
	Block	,034	1	,855
	Model	54,535	4	,000

Figura 7.3. *Testarea eficienței modelului înainte şi după includerea variabilei stresul perceput*
SPSS este o marcă înregistrată

Constatăm că şi al doilea model este semnificativ statistic, $G^2(4) = 54,535$, ca urmare, cel puțin una dintre VI incluse influențează evoluția VD. Se pune problema comparării celor două modele pentru a vedea care dintre ele este mai bun în explicarea apariției malformațiilor congenitale.

χ^2 al pasului apare notat ca Step χ^2, valoarea sa fiind $\Delta\chi^2(1) = 0.034$, $p > .854$. Valoarea este nesemnificativă statistic, ceea ce înseamnă că introducerea variabilei „stresul perceput" nu a adus îmbunătățiri suficiente modelului. Formula de calcul este foarte simplă:

$$\text{Step } \chi^2 = G^2 \text{ Model 2} - G^2 \text{ Model 1} \qquad (7.7)$$

unde modelul 2 este întotdeauna modelul cu mai mulți predictori incluşi.

Putem calcula şi diferența absolută dintre valorile BIC' ale celor două modele:

BIC'1 = -54,501 * 3*ln N = - 40,01
BIC'2 = -54,535 * 4*ln N = - 35,22
|BIC'1-BIC'2| = 4,79

Se observă că primul model este preferabil, dar consultând tabelul 7.1. observăm doar o probabilitate între .75 şi .95 ca acest model să fie mai bun decât al doilea.

Trebuie să privim întotdeauna şi dintr-o perspectivă teoretică atunci când comparăm două modele. Tehnicile statistice prezentate tind să prefere modelele ce dau dovadă de parcimonie în dauna altora mai complexe. Dacă există numeroase argumente care să ne facă să credem că o anume VI trebuie păstrată, în pofida aparențelor matematice, putem face acest lucru. În cazul de față este vorba de variabila stresul perceput. Deşi este cunoscută importanța ei în acest domeniu, ea nu a îmbunătățit primul model. Decizia rămâne la latitudinea cercetătorului, dacă va păstra sau nu variabila respectivă în model, chiar dacă includerea ei nu aduce un plus explicativ. De obicei, în regresiile cu scop predictiv, asemenea variabile sunt eliminate, deoarece obiectivul major este de a estima cât mai corect probabilitatea subiecților de a răspunde 1 sau 0 la criteriu.

7.3.3. Stabilirea eficienței explicative a modelului ales

Modelele pot trece pragul testării ipotezei statistice mai ales în cazul unor eşantioane foarte mari. Există însă o serie de indicatori, cu rol similar lui R^2 din regresia liniară, care arată în ce măsură VI influențează variația VD, indicând astfel, mărimea efectului.

Vom prezenta cinci asemenea indicatori: Pseudo - R^2, R^2 al lui McFadden, R^2 al lui Cox şi Snell, R^2 al lui Nagelkerke şi R^2 de stabilire a proporțiilor. Ultimii trei indicatori sunt calculați şi de SPSS.

Dacă R^2 din regresia liniară are rolul de a explica ce procent din dispersia VD este explicat de modificările VI, indicatorii similari din regresia logistică explică modificarea probabilității VD de a fi 0 sau 1 în funcție de variația VI. Aceşti indicatori nu testează o anumită ipoteză nulă, ci doar evaluează gradul de asociere între VI şi VD.

Pseudo-R^2, coeficient dezvoltat de Aldrich şi Nelson (apud Hair şi colab., 1998), are valori pozitive, subunitare. Are dezavantajul că valoarea sa maximă nu poate atinge 1. În schimb, tehnica poate fi folosită atât în cazul regresiei logistice binare, clasice, cât şi în cazul regresiei logistice multinominale sau a celei ordinale. Pentru a-l calcula utilizăm formula:

$$\text{Pseudo - } R^2 = 1 - (-2LL\text{model} / -2LL\text{null}) \qquad (7.8)$$

Se poate calcula şi un Pseudo-R^2 ajustat, pentru a ține seama de mărimea eşantionului (N) şi numărul de regresori din model (df '), excluzând interceptul:

$$Pseudo\_R^2 ajustat = \left(Pseudo\_R^2 - \frac{df'}{N-1} \right) * \left(\frac{N-1}{N-df'-1} \right) \qquad (7.9)$$

R^2 al lui McFadden (Long, 1997) este un indicator similar. Diferența este că McFadden ia în considerare log likelihood (LL) şi nu funcția –2LL. McFadden oferă posibilitatea unei ajustări (în engleză „Adjusted McFadden R^2"), pentru a nu spori artificial valoarea indicatorului o dată cu creşterea numărului de VI incluse în model.

$$R^2 \text{ McFadden} = 1 - (LL\text{model} / LL\text{null}) \qquad (7.10)$$

$$R^2 \text{ McFadden' (ajustat)} = 1 - [(LL\text{model} - df') / LL\text{null}] \qquad (7.11)$$

unde df ' este dat de numărul de parametri (VI) incluşi în model, exceptând interceptul.

R^2 al lui Cox şi Snell, respectiv **R^2 al lui Nagelkerke** (Garson, 2000) sunt indicatori prezentați în outputul SPSS-ului. Al doilea indicator se bazează pe

primul, fiind o ajustare a acestuia pentru ca R^2 să poată lua valori între 0 şi 1, asemeni lui R^2 din regresia liniară. Dintre toți indicatorii mai sus prezentați, R^2 al lui Negelkerke este singurul care poate lua valori între 0 şi 1, ceilalți având o valoare maximă subunitară.

R^2 al proporțiilor implică realizarea unui tabel de clasificare. Acesta conține, în cazul regresiei logistice dihotomice, 2 linii şi 2 coloane şi clasifică observațiile şi predicțiile în două grupe 0 sau 1 (ex. angajat sau şomer).

Tabelul 7.2. Exemplu de clasificare a observațiilor în funcție de datele culese şi de cele estimate

	Predicții şomeri (0)	Predicții angajați (1)
Observații şomeri (0)	32	5
Observații angajați (1)	3	60

O ecuație logistică va putea fi reprezentată printr-un asemenea tabel. Dacă VI sunt adecvate pentru explicarea VD, se va observa o similitudine între observații (datele actuale) şi predicții (pe baza VI) în a determina dacă o persoană are şanse mai mari să fie angajată sau, dimpotrivă, să fie şomeră.

Vom prezenta, pe rând, valorile indicatorilor discutați pentru exemplul nou-născuților, vizând modelul care nu include variabila "stresul perceput".

Pentru calculul lui Pseudo - R^2 trebuie să cunoaştem valorile abaterii modelului (-2LL model) şi ale abaterii inițiale (-2LL nul), iar pentru calculul lui R^2 a lui McFadden avem nevoie de valorile logaritmice pentru cele două situații (LL – loglikelihood). Dacă în variante mai vechi, precum SPSS 7.5., programul oferă valorile tuturor acestor abateri, versiunile mai noi, precum SPSS 11.0., permit calcularea acestor indicatori într-un mod indirect.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	109,970	,353	,483

Figura 7.4. *Indicatori de bază ai mărimii efectului în regresia logistică*
SPSS este o marcă înregistrată.

SPSS-ul oferă valorile G^2 şi -2LL pentru model. Se ştie însă că G^2 = -2LL null – (-2LL model), fapt ce ne ajută la calcularea abaterii inițiale. Astfel, dacă G^2 a fost 54,50, iar abaterea modelului a fost 109,97, rezultă că abaterea inițială (-2LL null) este 164,47.

Apoi, înlocuind în formulele prezentate mai sus (7.8; 7.9; 7.10 şi 7.11), vom obține:

Pseudo - R^2 = .3313 Pseudo - R^2 ajustat = .3222

R^2 a lui McFadden = .3313 R^2 a lui McFadden ajustat = .2948

Aceste valori pot fi interpretate în termeni de procente. În consecință, putem afirma că predictorii incluşi în model explică aproximativ între 29,48% şi 33,13% din variația probabilității VD de a fi 1 sau 0 (de a naşte un copil sănătos sau unul cu malformații).

R^2 al lui Cox şi Snell şi R^2 al lui Nagelkerke sunt indicatori asemănători ca mod de funcționare şi interpretare, fiind prezentați direct în SPSS (vezi figura 7.4.).

Se observă că valorile obținute sunt mai mari, mai ales în cazul lui R^2 al lui Nagelkerke. Conform lor, VI incluse în model explică aproximativ între 35,3%-48,3% din variația probabilității VD de a fi 1 sau 0.

R^2 de stabilire a proporțiilor este ultimul indicator prezentat. El măsoară gradul în care există o compatibilitate între observații şi predicții, adică gradul de corectitudine a predicțiilor. În constituirea sa are loc următorul demers. Pentru fiecare subiect în parte se calculează probabilitatea de a se situa într-una dintre cele două situații posibile „nou-născut cu malformații genetice" sau „nou-născut sănătos". Pe baza probabilităților calculate, subiecții sunt clasificați într-una dintre cele două grupe. După această clasificare a subiecților pe baza estimărilor obținute prin ecuația de regresie logistică, se compară corectitudinea clasificării cu rezultatele observate. În acest sens, există un rezultat global dat de raportul dintre predicțiile corecte şi totalul predicțiilor. Pentru exemplul dat, inspectând datele observate în figura 7.5, am obținut 77,6% de răspunsuri corecte (notat "overall percentage"):

R^2 proporții = (31 + 66) / (31 + 15 + 13 + 66) = 77,6%

Classification Table[a]

Observed			Predicted		
			HANDICAP		
			copii cu probleme	copii sanatosi	Percentage Correct
Step 1	HANDICAP	copii cu probleme	31	15	67,4
		copii sanatosi	13	66	83,5
	Overall Percentage				77,6

a. The cut value is ,500

Figura 7.5. *Clasificarea subiecților pe baza datelor observate şi a celor estimate*

SPSS este o marcă înregistrată.

Se pot calcula alți trei indicatori specifici (Wright, 1995). Opusă lui R^2 de stabilire a proporțiilor, este valoarea predicțiilor negative dată de suma predicțiilor greşite supra totalul predicțiilor. Ea este egală cu 1 - R^2 de stabilire a proporțiilor, în

acest caz 23,4%. Astfel, aproximativ un caz din patru va fi prezis incorect de model. În schimb, acesta se va dovedi a fi valid pentru celelalte ¾ din cazuri.

De asemenea, există gradul de sensibilitate (în engleză „sensitivity") care constă în raportul dintre numărul de predicții corecte şi numărul de observații ale evenimentului 1. Pentru exemplul dat, este vorba de raportul dintre numărul de predicții corecte referitoare la răspunsul „nou-născut sănătos", din numărul total de copii sănătoşi observați. Vom obține 66 / (66 + 13), adică 83,50%. În acest context se poate vorbi de 13 cazuri de diagnosticare greşită (falşi negativi – prezişi cu malformații, dar născuți sănătoşi). Similar, se poate calcula şi specificitatea (în engleză „specificity"), care se referă la raportul dintre numărul de predicții corecte şi numărul de observații ale evenimentului 0. Pentru situația dată au fost estimați corect, pe baza modelului, 31 dintre cei 46 de nou-născuți cu malformații genetice, adică 67,40% [31 / (31 + 15)]. Se observă un număr de 15 cazuri diagnosticate greşit (falşi pozitivi – prezişi sănătoşi, dar născuți cu unele malformații).

Interpretarea se face ținând cont atât de valoarea globală a lui R^2 de stabilire a proporțiilor, cât şi în funcție de sensibilitate şi specificitate. Pentru exemplul nostru am putea spune că modelul propus reuşeşte să ofere o mai bună predicție a cazurilor de copii născuți sănătoşi (în proporție de 83,84% din totalul copiilor născuți sănătoşi), comparativ cu predicția naşterii unor copii cu malformații (numai 65,26% din cazurile de copii născuți cu diverse handicapuri). Totuşi, valorile obținute sunt situate mult peste media unui răspuns dat la întâmplare (50% din cazuri).

Trebuie să precizăm că împărțirea subiecților într-una dintre cele două grupe, pe baza rezultatelor estimate, are ca valoare standard în SPSS probabilitatea de .50. Dacă un subiect are o probabilitate estimată pentru evenimentul 1 („nou-născuți sănătoşi") mai mare de .50 se consideră că va fi un nou-născut fără probleme. Dacă probabilitatea pentru acelaşi răspuns va fi mai mică de .50, se va estima că va fi un nou-născut cu malformații genetice. SPSS-ul oferă posibilitatea de a modifica această valoare critică (în engleză „cut value"), măsură ce va conduce la modificări ale valorii R^2 al proporțiilor sau ale indicatorilor subsidiari acestuia.

De exemplu, schimbarea valorii critice de clasificare de la probabilitatea de .50 la cea de .70 ca subiecții să fie încadrați ca „nou-născuți sănătoşi", va conduce la următorul tabel de clasificare.

Classification Table[a]

Observed			Predicted		
			HANDICAP		
			copii cu probleme	copii sanatosi	Percentage Correct
Step 1	HANDICAP	copii cu probleme	42	4	91,3
		copii sanatosi	21	58	73,4
	Overall Percentage				80,0

a. The cut value is ,700

Figura 7.6. *Reclasificarea subiecților pe baza datelor observate şi a celor estimate*
SPSS este o marcă înregistrată.

În acest caz, se observă că ridicarea probabilității pentru a încadra nou-născuții în categoria celor sănătoşi duce la o mai bună estimare a viitorilor nou-născuți care vor avea malformații genetice, 91,3% dintre aceştia fiind identificați corect. În schimb, scade la 73,4% puterea de estimare corectă a nou-născuților estimați ca fiind sănătoşi.

În aceste condiții, decizia de a alege punctul critic de clasificare ține exclusiv de cercetător, pe baza obiectivelor urmărite. Astfel, în cazul când este mai important să identificăm corect, încă din primele luni de sarcină, dacă copiii vor fi sănătoşi, vom alege un punct critic care să ne ajute să estimăm cât mai corect acest lucru. Dacă, în schimb, este mai important să identificăm de timpuriu copiii cu malformații genetice, pentru a permite avortul, atunci criteriul ales va trebui să faciliteze detectarea corectă a acestora.

De asemenea, trebuie să ținem seama în SPSS de modul de codare a VD. Dacă codurile date celor două răspunsuri posibile sunt 0 şi 1, programul va păstra acest sistem de clasificare. Dacă însă, cele două răspunsuri vor fi codate 1 şi 2, SPSS-ul va recoda automat valorile VD în 0 şi 1. Acest lucru s-a întâmplat şi în cazul explicat. Datele VD au fost introduse inițial cu următoarele coduri: 1- copii cu malformații genetice; 2- copii sănătoşi. Acestea au fost recodate automat în: 0 – copii cu malformații genetice şi 1- copii sănătoşi. Această schimbare poate deruta procesul de interpretare a coeficienților logit semnificativi, de aceea se recomandă ca întotdeauna VD din regresia logistică să fie codată, de la început, cu valorile 1 sau 0.

7.3.4. Testarea şi interpretarea coeficienților logit

Vom defini câteva concepte pentru a înțelege mai bine această secțiune.

Coeficienții logit din regresia logistică corespund coeficienților b din regresia liniară. Sunt utilizați în ecuația regresiei logistice pentru a estima raportul de şanse ca variabila dihotomică să fie 1 sau 0. Mai exact, ei indică măsura în care

se modifică şansele ca VD să fie 1 sau 0 când predictorul asociat creşte cu o unitate, iar ceilalți predictori rămân constanți.

Coeficienții logit trebuie testați pentru a vedea dacă au o influență semnificativă asupra VD. Există în special două tehnici folosite pentru acest scop: testul Wald şi testul χ^2 al pasului.

Testul statistic Wald este o metodă utilizată frecvent în regresia logistică pentru a testa ipoteza nulă conform căreia un anumit coeficient logit este zero. Cu alte cuvinte, testul Wald ne arată dacă predictorul respectiv influențează sau nu criteriul. Rezultatului obținut îi corespunde o valoare t sau z, similară ca interpretare cu situația din regresia liniară. În acest mod, pot fi eliminate din model acele VI care nu au fost semnificative. Totuşi prin testul Wald există posibilitatea comiterii unei erori de tip II în cazul unor coeficienții logit foarte mari, şi implicit, a unor erori standard ridicate. Aceasta semnifică respingerea unei VI, atunci când ea este semnificativă. În această situație (ex. un venit exprimat în milioane lei) sau în cazul variabilelor dummy, ar fi mai bine să se utilizeze o altă tehnică pentru a reduce şansele de a comite eroarea de tip II. Alternativa implică calcularea lui χ^2 **al pasului**, procedeu descris anterior, după introducerea variabilelor în cauză. Diferența obținută este evaluată ca χ^2 cu k-1 grade de libertate, k reprezentând numărul de categorii ale variabilei introduse. Dacă rezultatul este semnificativ statistic la $p < .05$ vom păstra variabila respectivă sau blocul de variabile dummy. În cazul în care nu sesizăm diferențe semnificative între modelul explicativ care include variabila în cauză şi cel care o exclude, putem elimina variabila respectivă pentru a simplifica modelul (Garson, 2000).

Coeficienții logit obținuți, deşi semnificativi statistic, nu pot fi comparați direct între ei din cauza problemelor legate de utilizarea unor scale de măsurare distincte. Pentru a rezolva această problemă există posibilitatea de a calcula un coeficient logit standardizat, indice similar lui beta din regresia liniară, sau de a estima contribuția parțială a lui R pentru fiecare VI în parte.

Coeficienții logit standardizați corespund coeficienților beta din regresia liniară. Sunt utilizați pentru a compara influența diferiților predictori asupra criteriului, în situația în care cele două VI sunt măsurate pe scale diferite (ex. vârsta evaluată în ani, venitul exprimat în lei/lună). De notat că pentru variabilele de tip dummy, coeficienții standardizați nu au nici o semnificație inteligibilă. SPSS-ul nu oferă coeficienții logit standardizați, însă o măsură acceptabilă este multiplicarea fiecărui coeficient logit cu abaterea standard a variabilei respective. Rezultatele obținute în valori absolute sunt apoi ordonate pe ranguri. O altă posibilitate constă în standardizarea tuturor VI.

Contribuția parțială a lui R este o metodă de stabilire a importanței fiecărei VI incluse în model fără a fi influențat de scala de măsurare. R parțial determină ce procent din dispersie se datorează fiecărui coeficient logit semnificativ în parte când ceilalți sunt ținuți la valori constante. Valoarea este dependentă de testul Wald şi de eroarea standard, fiind astfel supusă neajunsurilor acestora amintite

anterior însă are avantajul de a fi prezentat direct în variantele mai vechi ale SPSS-ului.

Pentru situațiile descrise mai sus, în care există riscul ca testul Wald să ofere valori eronate, este mai utilă calcularea coeficienților standardizați logit sau multiplicarea coeficienților logit cu abaterea standard a VI respective (Garson, 2000).

Din punctul de vedere al interpretării coeficienților logit, valorile negative indică o descreştere a şanselor de a obține evenimentul codat 1, pe măsură ce valoarea predictorului creşte. În schimb, un coeficient logit pozitiv indică o relație directă, şansele crescând o dată cu mărirea valorii predictorului. Pe baza coeficienților logit, prezentați în figura 7.7, se poate scrie ecuația logistică a modelului propus.

ln Handicap = 4,57+1,36*Fără alcool–2,23*Factori genetici–0,09*Vârsta mamei

Dacă coeficienții logit obținuți sunt semnificativi statistic, următorul pas constă în interpretarea lor. Vom regăsi două modalități principale de interpretare: (a) **interpretarea prin factori** şi (b) **interpretarea prin procente** (Long, 1997). Din figura 7.7 se observă că toți cei trei predictori exercită o influență semnificativă din punct de vedere statistic asupra evoluției probabilității criteriului de a lua valoarea 1, adică de a se naşte copii sănătoşi. De exemplu, predictorul vârsta mamei la naştere are un coeficient logit de -0,09, rezultat semnificativ statistic Wald (1) = 10,11, p < .01.

Variable	B	S.E.	Wald	df	Sig	R
ALCOOL(1)	1,3671	,6142	4,9541	1	,0260	,1340
FACT. GENETIC(1)	-2,2370	,5111	19,1566	1	,0000	-,3230
VARSTA	-,0938	,0295	10,1136	1	,0015	-,2221
Constant	4,5707	1,1631	15,4425	1	,0001	

		95% CI for Exp(B)	
Variable	Exp(B)	Lower	Upper
ALCOOL(1)	3,9238	1,1774	13,0768
FACT. GENETICI(1)	,1068	,0392	,2908
VARSTA	,9105	,8593	,9647

Figura 7.7. *Output oferit cu privire la valorile coeficienților logit*
SPSS este o marcă înregistrată.

În cazul celorlalți doi predictori care sunt variabile dummy, testul Wald, deşi nu este cel mai potrivit de a fi aplicat în asemenea cazuri, indică, de asemenea, coeficienți logit semnificativi statistic. Acest lucru este confirmat dacă se

calculează, pe rând, câte un χ^2 al pasului pentru modele care exclud succesiv aceste variabile dummy.

Înainte de a explica cele două modalități de interpretare, prin factori, respectiv prin procente, se cuvine să precizăm cum au fost codate toate aceste variabile pentru a facilita înțelegerea acestui demers. Astfel, vârsta a fost exprimată în ani, iar VD a avut doar două coduri asociate: 1 – nou-născuți sănătoşi şi 0 – nou-născuți cu malformații genetice (vezi, de asemenea, figura 7.8).

În utilizarea SPSS-ului trebuie să se diferențieze variabilele categoriale de cele continue (Garson, 2000). Pentru aceasta activăm opțiunea "Categorical" în cadrul ferestrei legate de regresia logistică. Se va deschide o nouă fereastră ce conține o listă de "Covariates" cu toate VI incluse. Vom selecta din listă doar predictorii care trebuie transformați în variabile dummy, adică doar variabilele măsurate prin scale nominale, şi îi vom muta în lista denumită "Categorical Covariates". Opțiunea standard a SPSS-ului este de a transforma ultima categorie de răspuns codată ("last") în valoare de referință. Prin acest procedeu, SPSS-ul permite crearea automată a variabilelor dummy. Modul de codare a variabilelor dummy din exemplu apare în output-ul SPSS-ului (figura 7.8).

De exemplu, din totalul de 125 de mame chestionate, 42 au consumat alcool regulat pe parcursul sarcinii, iar 83 s-au abținut de la acest lucru. Datele au fost introduse în calculator cu următoarele coduri: 1 – dacă mama nu a consumat alcool în timpul sarcinii; 2 – dacă mama a consumat alcool în timpul sarcinii. Prin atenționarea SPSS-ului că predictorul respectiv necesită transformarea în variabilă dummy, 1 a devenit codul pentru mamele care nu consumă alcool, iar 0 – codul pentru mamele care au consumat alcool regulat în acea perioadă.

Parameter	Value	Freq	Coding (1)
FACTORI GENETICI			
Malformații în familie	1	62	1,000
Lipsa de malformatii	2	63	,000
ALCOOL			
nu bea in timpul sarcinii	1	42	1,000
bea in timpul sarcinii	2	83	,000

Figura 7.8. *Output SPSS 11.0 cu privire la modul de transformare a variabilelor nominale în variabile dummy*
SPSS este o marcă înregistrată.

(a) *Interpretarea prin factori* implică valorile exponentului coeficientului logit obținut – exp(b) (vezi figura 7.7). De exemplu, în cazul variabilei consum de alcool, b a fost 1,36, iar factorul rezultat, exp(b) a fost 3,92. Pe baza acestor rezultate putem afirma că persoanele care nu au consumat alcool au de patru ori (3,92 ori) mai multe şanse de a naşte copii sănătoşi în comparație cu o persoană care a

consumat alcool în timpul sarcinii, în condițiile în care vom păstra constante valorile celorlalte VI implicate în model.

Pentru variabila vârstă avem exp (-0,09) = 0,9139. Fiecare an de viață în plus modifică cu un factor de 0,91 şansele de a naşte un copil sănătos. Valorile subunitare ale factorului înseamnă de fapt o relație inversă între cele două elemente (ex. mai mulți ani, mai puține şanse de a naşte un copil sănătos).

Formula generală utilizată în interpretarea prin factori afirmă că avem o creştere (valori supraunitare) sau o descreştere (valori subunitare) de exp(b) ori în ceea ce priveşte şansele ca VD să fie 1 când VI respectivă creşte cu o unitate, iar ceilalți predictori sunt menținuți la valori constante.

Pentru VI continue putem interpreta valorile b şi pentru situații când VI creşte mai mult de o unitate. Vom exemplifica cum se modifică şansele de a naşte un copil sănătos la interval de 10 ani diferență. Formula de calcul devine: exp (b*10) = exp (-0,09*10) = 0,4067. Astfel, o femeie de 30 de ani are 0,4067 şanse (mai puțin de jumătate) de a naşte un copil sănătos în comparație cu o alta de 20 de ani. Această interpretare este valabilă dacă ceilalți predictori iau valori similare, de exemplu, fie ambele femei comparate nu au consumat alcool în perioda sarcinii, fie au fost comparate două femei care au consumat alcool în perioada sarcinii.

(b) Cea de-a doua posibilitate de interpretare este cea *prin procente*, caz în care se utilizează formula:

$$100\left(e^{b}-1\right) \qquad (7.12)$$

Pentru variabila „consum de alcool" vom obține 100 (3,925-1), adică 292,5%. O valoare negativă indică o descreştere în şansele ca VD să fie 1 atunci când VI creşte cu o unitate, în timp ce o valoare pozitivă indică o creştere a lor.

Astfel, abstinența de la consumul de alcool în perioada sarcinii creşte cu 292,5% şansele de a naşte un copil sănătos comparativ cu o altă persoană ce a consumat alcool în acea perioadă, păstrând toate celelalte variabile constante.

Pentru variabila vârstă, vom obține 100 (0,91-1), adică –8,95%. Aceasta înseamnă că fiecare an de viață în plus determină o descreştere cu 8,95% a şansei de a da naştere unui copil sănătos, păstrând toate celelalte variabile constante.

7.3.5. Stabilirea unor profile ideale

Pe baza ecuației logistice putem stabili profile pentru diverse valori ale VI. Avantajul acestora constă în posibilitatea de a manipula mai multe VI, în acelaşi timp, pentru a observa modificările în evoluția VD (Freese, 2000). Reamintim că ecuația de regresie pentru exemplul nostru este:

ln Handicap = 4,57 + 1,36*Alcool – 2,23*Factori genetici – 0,09*Vârsta la naştere

Vom compara două persoane diferite din perspectiva probabilității ca VD să fie 1 (nou-născut sănătos):

a) o mamă de 20 de ani, care nu a consumat alcool în perioada sarcinii şi care nu are între rudele de gradul I şi II antecedente de copii cu handicap;
b) o mamă de 40 de ani, care a consumat alcool în timpul sarcinii şi care are în familie persoane cu malformații congenitale.

VD (a) = exp (4,57 – 0,09*20 + 1,36*1 – 2,23*0) = 4,13 (a)
VD (b) = exp (4,57 – 0,09*40 + 1,36*0 – 2,23*1) = - 1,26 (b)

Acestea sunt apoi transformate în probabilități pe baza formulei 7.4. Vom obține o probabilitate de 0,9837 pentru cazul a şi una de 0,2205 pentru cazul b. O probabilitate de 0,50 înseamnă şanse egale pentru ambele categorii VD. Observăm că primul subiect este mai probabil să nască un copil sănătos, iar al doilea, unul cu malformații genetice.

Pe baza probabilității ca VD să fie 1 putem calcula şi probabilitatea ca VD să fie 0 (q = 1-p). Astfel, probabilitatea de a naşte un copil cu malformații genetice este de 0,0163 în cazul primei mame şi de 0,7795, în cel de-al doilea caz. În ce priveşte şansele, observăm că şansa de a naşte un copil sănătos este în primul caz de 60,35 ori mai mare (.9837/.0163) în comparație cu naşterea unui copil cu malformații genetice. În schimb, există de 3,54 ori mai puține şanse de a naşte un copil sănătos pentru cel de-al doilea caz (.2205 / .7795 = 0,282). De asemenea, putem compara direct rezultatele celor două persoane. Astfel, raportul de şanse arată că prima persoană are 213,40 ori mai multe şanse de a naşte un copil sănătos în comparație cu cea de-a doua persoană [(.9837/.0163) / (.2205/.7795)].

7.3.6. Analiza cazurilor influente asupra modelului

SPSS oferă, prin activarea butonului "Save" din regresia logistică, posibilitatea de a calcula indicatori ce descriu sau analizează fiecare caz în parte, asemenea situației amintite în cazul regresiei liniare.

De exemplu, prin bifarea "Predicted Values" putem stabili probabilitatea VD de a fi 1 pentru fiecare observație sau putem determina grupul în care a fost inclus cazul respectiv (Hair et al., 1998).

Prin opțiunea "Influence" avem posibilitatea de a măsura influența fiecărui caz asupra modelului. În tabelul 7.3. vor fi descrişi principalii indicatori utilizați în detectarea cazurilor influente.

Cu ajutorul acestor indicatori putem studia cazurile care nu se încadrează bine în modelul explicativ propus. Cu alte cuvinte, putem identifica atât cazurile neobişnuite, cât şi cele care au contribuit cel mai mult la obținerea valorilor logit din ecuația de regresie. Aceste informații ne ajută în formarea unei impresii generale despre stabilitatea soluției găsite, adică despre gradul de specificitate al soluției la lotul de subiecți testat. Acesta are implicații deosebite, mai ales în regresia cu scop

predictiv. Ele sunt legate de posibilitatea de a generaliza rezultatele pe alte loturi de subiecți. Ca regulă generală, cu cât există mai multe cazuri neobişnuite sau influente, cu atât sunt mai mici şansele de generalizare a rezultatelor la nivelul populației din care a fost extras lotul testat.

Tabelul 7.3. Principalii indicatori de detectare a cazurilor influente

Denumire -SPSS	Descriere	Valori critice
Cook's Distance	Evaluează în ce măsură eliminarea cazului respectiv schimbă probabilitățile ca VD să fie 1 pentru celelalte observații.	> 4 / (n-k-1)
DfBeta	Măsoară modificările coeficientului logit pentru fiecare VI în parte, în situația excluderii cazului respectiv. Aceste valori se standardizează apoi.	Depăşirea intervalului ±2
"Leverage value"	Evaluează măsura în care un caz influențează abaterea modelului şi implicit testul G^2.	> 3k/n

unde n este volumul eşantionului, iar k numărul de VI.

7.4. Exemplu de regresie logistică binară – cazul infidelității

În scrierea secțiunii de analiză a datelor în articolele care utilizează regresia logistică trebuie prezentate următoarele informații:

- precizarea într-un paragraf a variabilelor implicate în regresia logistică, atât a predictorilor (VI), cât şi a criteriului (VD), precum şi a justificării pentru alegerea acestei tehnici, pornind de la ipoteza sau obiectivul studiului;
- precizarea într-un alt paragraf a datelor generale cu privire la eficiența modelului testat, a puterii sale explicative şi a eventualelor comparații cu alte modele alternative;
- prezentarea unui tabel sumar al valorilor coeficienților logit şi a exponenților acestora, precum şi a indicatorilor utilizați pentru testarea gradului de semnificație a acestora;
- în cazul regresiilor în scop predictiv se poate prezenta şi tabelul cu R^2 al proporțiilor, pentru a afişa gradul de corectitudine a clasificării subiecților pe baza ecuației de regresie observate;
- în cazul regresiilor cu scop explicativ, ar fi util de trecut în tabelul cu coeficienții logit, valorile R sau coeficienții logit standardizați, pentru a vedea contribuția fiecărei variabile la explicarea variabilei independente; de asemenea, în cazul utilizării metodei ierarhice se va trece χ^2 al pasului, pentru fiecare treaptă adăugată;
- interpretarea coeficienților logit obținuți folosind, fie varianta factorilor, fie pe cea a procentelor din analiza raportului şanselor – exp(b).

Pentru exemplificare, vom prezenta secțiunea de analiză a datelor într-o cercetare ipotetică, cu un scop predictiv, acela de a găsi un set de variabile care să prezică cel mai bine probabilitatea unei persoane de a fi fidelă, respectiv de a fi infidelă în plan sexual, față de partenerul sau partenera sa. În cercetările anterioare au fost identificați câțiva factori care par a influența acest comportament. Ei pot fi împărțiți în factori personali (sexul persoanei, vârsta, nivelul de educație, venitul realizat, tipul de religie adoptată, starea civilă, statutul social), factori situaționali (distanța fizică dintre parteneri, numărul de parteneri de sex opus eligibili la locul de muncă, vârsta la care a avut loc căsătoria (prima căsătorie) sau factori psihologici (gradul de satisfacție maritală în cuplu, gradul de implicare în relație, nevoia de senzații tari, stilul de ataşament etc.) (Atkins, Baucom şi Jacobson, 2001; Bogaert şi Sadava, 2002; Drigotas şi Barta, 2001; Drigotas, Safstrom şi Gentilia, 1999; Zak şi colab., 2002).

Dintre aceştia, am decis să găsim nişte predictori care să estimeze cât mai corect persoanele care vor da dovadă de fidelitate în plan sexual, pe un lot ipotetic de 200 de studenți necăsătoriți, având vârsta cuprinsă între 18 şi 30 de ani.

Secțiunea de analiză a datelor ar putea lua următoarea formă:

Pentru a găsi un model predictiv care să estimeze cât mai corect fidelitatea sau infidelitatea în plan sexual, am utilizat ca predictori un număr de opt variabile: vârsta; sexul persoanei; comportamentul religios (frecvența la biserică; păstrarea ritualurilor în post etc.), atitudinea față de relații, stilul de ataşament cu cele trei scale (sigur; ambivalent-anxios; evitant) şi nevoia de senzații tari.

Utilizând metoda simultană, modelul testat are un $G^2(8) = 42.15$, $p < .01$, valoare semnificativă satistic, iar R^2 Nagelkerke a fost .56. De altfel, cinci predictori din opt exercită o influență semnificativă statistic asupra probabilității ca studenții să fie fideli, respectiv infideli. Ținând cont de faptul că scopul acestei cercetări este exclusiv predictiv, precum şi de faptul că modelul predictiv redus, ce include doar cele cinci variabile semnificative statistic, are aproape aceeaşi putere explicativă, R^2 Nagelkerke fiind .53, iar $\Delta G^2(3) = 3.00$, $p > .05$, vom interpreta acest model restrâns, în care $G^2(5) = 39.15$, $p < .01$. Cei trei predictori eliminați sunt scorurile subiecților la stilul de ataşament sigur şi la cel de evitare, respectiv vârsta, deoarece coeficienții logit asociați acestora nu au fost semnificativi statistic.

Tabel cu clasificarea subiecților în funcție de valorile observate şi de cele estimate

		Rezultate estimate	
		Studenți infideli	Studenți fideli
Rezultate observate	Studenți infideli	31	2
	Studenți fideli	20	147

Acest model cu cinci predictori reuşeşte să estimeze corect 89% din cazuri, un procent uşor superior fiind în identificarea cazurilor de infidelitate (94%,

adică 31 din cele 33 de cazuri), decât în identificarea cazurilor de fidelitate (88%, adică 147 din cele 167 de observații).

O asemenea eficiență predictivă este îmbucurătoare, aproximativ 9 din 10 studenți fiind identificați corect ca aparținând, fie categoriei celor fideli sexual, fie celor infideli sexual. Mai mult, doar 2 dintre studenții care au fost infideli nu au fost identificați corect prin ecuația de regresie rezultată, ceilalți 31 fiind identificați corect.

Tabelul cu coeficienții logit şi raportul şanselor pentru fiecare predictor din model

Predictori	b	S.E.	Exp(b)	Wald
Sexul persoanei	0,97	0,33	2,64	7,23 **
Comportamentul religios	- 0,33	0,14	0,72	4,91 *
Atitudinea față de relație	0,61	0,17	1,84	10,25 **
Ataşament ambivalent-anxios	0,22	0,10	1,25	5,10 *
Nevoia de senzații tari	0,30	0,19	1,35	6,32 **
Constanta (Interceptul)	0,25	0,06		12,46 **

* p < .05, ** p < .01

Se constată că persoanele de sex masculin au de 2,64 ori mai multe şanse de a fi infidele față de partenerele (partenerii) lor, în comparație cu persoanele de sex feminin, în condițiile în care păstrăm constante celelalte variabile. De asemenea, şansele ca persoanele care adoptă un comportament religios (variabilă măsurată pe o scală cu 10 puncte) să adopte un comportament infidel scad cu 0,72, dacă scorul la comportamentul religios creşte cu un punct, în aceleaşi condiții, de păstrare a celorlalte variabile la valori constante. În schimb, şansele de infidelitate în relație aproape se dublează (b = 1,84) cu fiecare punct adăugat la variabila atitudinea față de relație, unde valorile ridicate indică o atitudine liberală. Şansele cresc de 1,25 ori, cu fiecare punct adăugat în scala de ataşament de tip ambivalent-anxios. Cu alte cuvinte, un asemenea stil de ataşament favorizează apariția infidelității în plan sexual. În sfârşit, nevoia de senzații tari favorizează, de asemenea, comportamentul infidel. Cu fiecare punct în plus la scorul acestui predictor, şansele de a fi infidel cresc de 1,35 ori, în condițiile menținerii la valori constante a celorlalte variabile.

Prin urmare, profilul persoanelor cu risc ridicat de a se angaja în relații sexuale în afara cuplului, rezultat din lotul de 200 de studenți testați, este acela al unui bărbat, care are un comportament religios redus (nu frecventează foarte des slujbele săptămânale, nu ține post etc.), care are o atitudine liberală, deschisă, față de cuplu (ex. consideră normal să participe la petreceri şi fără partener sau parteneră), care are un stil de ataşament de tip ambivalent-anxios şi o nevoie puternică de senzații tari.

BIBLIOGRAFIE

Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley & Sons.

Atkins, D.C., Baucom, D.H., Jacobson, N.S. (2001). Understanding infidelity: Correlates in a National Random Sample. *Journal of Family Psychology*, 15(4), 735-749.

Bogaert, A.F, Sadava, S. (2002). Adult attachment and sexual behavior. *Personal Relationships*, 9, 191-204.

Drigotas, S.M., Safstrom, C.A., Gentilia, T. (1999). An investment model prediction of dating infidelity. *Journal of Personality and Social Psychology*, 77(3), 509-524.

Drigotas, S.M., Barta, W. (2001). The cheating heart: Scientific explorations of infidelity. *Current Directions in Psychological Science*, 10(5), 177-180.

Freese, J. (2000). *Categorical analysis*. Note de curs. ICPSR Summer Program in Quantitative Methods of Social Research, Ann Arbor.

Hair, J.E.Jr., Anderson, R.E., Tatham, R.L., Black, W.C. (1998). *Multivariate data analysis* (fifth edition). New Jersey, Prentice Hall.

Garson (2000). Găsit la 14.IX pe www2.chass.ncsu.edu/garson/pa765/logistic.htm

Long, Scott (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, Sage.

Menard, S. (1995). *Applied logistic regression analysis*. Publications Series: Quantitative Applications in the Social Sciences, nr.106, Thousand Oaks, Sage .

Wickens,T.D. (1998). Categorical data analysis. *Annual Review in Psychology*, 48, 537-558.

Wright, R.E. (1995). *Logistic Regression*. În L.G. Grimm şi P.R. Yarnold (eds.). Reading and understanding multivariate statistics (p. 217-244). Washington, DC: American Psychological Association.

Zak, A., Coulter, C., Giglio, S., Hall, J., Sanford, S., Pellowski, N. (2002). Do his friends and family like me? Predictors of infidelity in intimate relationships. *North American Journal of Psychology*, 4(2), 287-290.

CAPITOLUL 8

SCALAREA MULTIDIMENSIONALĂ

Se pare că Torgerson a fost primul care a propus termenul de scalare multidimensională (Notelaers, 2002). Acest termen a fost adoptat mai ales de cercetările din domeniul psihofizicii, în timp ce în domeniul psihometriei s-a utilizat mult timp termenul de "analiza spațiului celui mai mic" (în engleză „smallest space analysis"). Chiar dacă au existat unele distincții între aceşti termeni, în special la nivelul algoritmilor de calcul, principiul este acelaşi. Prin urmare, vom utiliza, pentru ambele situații, termenul generic de scalare multidimensională.

În cadrul scalării multidimensionale se regăsesc un grup de tehnici menite să identifice gradul de proximitate perceptivă între diferite aspecte studiate. Principiul după care funcționează aceste tehnici este simplu: există o anumită distanță psihologică între obiecte, iar această distanță poate fi reprezentată grafic sub forma unei hărți perceptive. Astfel, gradul de similaritate / disimilaritate între obiecte (ex. concepte, trăsături comportamentale, imagini, afecte) este transpus grafic sub forma unei distanțe între puncte, acestea nefiind altceva decât proiecții ale obiectelor în plan sau în spațiu. Cu cât două obiecte sunt percepute mai diferit, cu atât distanța dintre punctele reprezentate grafic va fi mai mare. În schimb, dacă două obiecte vor fi percepute ca fiind asemănătoare, distanța dintre punctele corespunzătoare va fi mai mică. Prin acest demers, MDS (multidimensional scaling) se constituie ca o tehnică multivariată, exploratorie, ce poate fi utilizată pentru a vizualiza proximitățile dintre obiecte într-un spațiu redus dimensional. Dimensiunile (axele) acestui spațiu vor constitui criteriile (variabilele) latente care stau la baza stabilirii proximității dintre obiecte. Din acest punct de vedere, scalarea multidimensională se apropie de analiza factorială.

Analiza factorială presupune reducerea unui număr mai mare de variabile la câteva variabile latente, denumite factori. În acelaşi mod, în MDS se încearcă reprezentarea grafică în cât mai puține dimensiuni (axe) a unui număr mai mare de puncte (obiecte). Ambele metode permit realizarea unor rotiri ortogonale. Există, însă, diferențe importante între cele două tehnici. Dacă analiza factorială, exceptând tehnica Q, necesită un număr relativ mare de răspunsuri (subiecți) şi un număr relativ mic de variabile ce trebuie factorizate, scalarea multidimensională necesită, mai degrabă, un număr mare de puncte (variabile) şi poate fi desfăşurată, în anumite cazuri, chiar pe un singur subiect.

8.1. Modele de scalare multidimensională

Din punct de vedere matematic, problema principală în MDS se reduce la găsirea unei modalități de transformare a gradului de similaritate / disimilaritate perceptivă dintre obiecte în distanțe între puncte reprezentate grafic. În acest scop a fost creată o serie de algoritmi de calcul, a căror selecție este dependentă de câțiva factori:

a) tipul de date prin care au fost evaluate proximitățile psihologice dintre obiecte;
b) sursa provenienței datelor: un singur grup, mai multe grupe, o singură persoană;
c) tipul de relații între obiectele avute în vedere.

În funcție de aceste aspecte, există mai multe categorii de modele MDS (van Deun şi Delbeke, 2000):

a) modele MDS metrice vs. non-metrice

Conform primului criteriu, cel al tipului de date analizate, proximitățile obținute prin scale de tip interval sau raport (proporții) necesită tehnici metrice, în timp ce datele obținute prin scale de tip ordinal sau nominal fac apel la tehnici non-metrice. Deşi primul gen de tehnici a fost inventat mai întâi, cele non-metrice sunt mult mai des întâlnite în literatura de specialitate (Stalans, 1998). Cel mai întâlnit tip de date utilizate în MDS sunt cele ordinale (ex. ierarhizări, scale Likert). De pildă, pentru a compara gradul de proximitate dintre două direcții ale psihologiei: psihologia cognitivă şi behaviorism, se poate utiliza o scală de comparație cu şapte trepte, unde "1" ar însemna "foarte asemănătoare", iar "7", "foarte diferite".

b) modele MDS sintetice vs. analitice

Primele (în engleză „unweighted models") pornesc de la premisa că toți subiecții atribuie aceeaşi importanță şi utilizează aceleaşi criterii de evaluare. Mai mult, obiectivul unui studiu, utilizând asemenea modele, este de a observa cum îşi reprezintă toți subiecții dintr-un eşantion relațiile de proximitate psihologică dintre obiecte. Pentru a lucra cu asemenea date se obişnuieşte să se calculeze media răspunsurilor provenite de la toți subiecții din eşantion pentru fiecare comparație făcută. Modelele analitice se împart în două categorii: modele replicative (în engleză „replicative models") şi modele diferențiale (în engleză „weighted models"). Cele dintâi se referă la situații în care se realizează comparații repetate la intervale diferite de timp, pe acelaşi lot de subiecți sau pe un singur subiect. Cele din urmă satisfac situațiile în care există o singură testare, dar în care suntem interesați de diferențele interindividuale sau intergrupale existente în cadrul unui eşantion. Din această ultimă perspectivă, cel mai probabil, vom constata diferențe între psihologia cognitivă şi behaviorism, dacă vom întreba specialişti în domeniul psihologiei transpersonale sau profesionişti în domeniul inteligenței artificiale.

c) modele de similaritate vs. modele unfolding (extinse)

Proximitatea, conform teoriei lui Coombs, poate conține atât relații de similaritate (gradul de apropiere între două oraşe), cât şi relații de alegere

preferențială (prefer oraşul A lui B). Primul tip de date face referire la modele clasice de scalare multidimensională, în timp ce alegerile preferențiale incumbă modele unfolding. Totuşi, acestea din urmă pot fi transformate în date de tip similaritate şi supuse modelelor de tip MDS (Rotariu şi colab., 1999).

Dincolo de aceste diferențieri, importante din perspectivă teoretică, există diferențe şi în utilizarea unor algoritmi de calcul. Aceasta se datorează, în mare parte, lipsei de unitate conceptuală în evoluția MDS şi se regăseşte în modalitățile diferite de transformare a proximității psihologice în distanțe între puncte reprezentate grafic, dar şi în diversitatea de programe statistice existente. Vom aminti pe scurt câțiva dintre algoritmi. Cititorul trebuie să rețină că alegerea corectă a programul statistic ține de tipul de algoritmi de care acesta dispune. Fiecărui design MDS îi poate corespunde rularea unuia sau a mai multor algoritmi. Dacă nu este selectat algoritmul MDS adecvat design-ului cercetării, se vor obține rezultate eronate.

Pornind de la aceste precizări, au fost dezvoltate mai multe proceduri de scalare multidimensională precum: KYST; MINISSA; MDPREF; TRISOSCAL; MULTISCALE; INDSCAL; ALSCAL; PROXSCAL. Aceste proceduri sunt cuprinse într-unul sau în mai multe programe statistice cum ar fi:

- MDS(X) Library of Computers programs (Coxon, 1982) este un program care conține majoritatea procedurilor amintite mai sus, alături de alte câteva şi poate fi descărcat de pe internet gratuit, cu licență pe o perioadă de 60 de zile. Adresa web este: http://www.newmdsx.com
- SPSS-ul conține în meniu trei dintre procedurile amintite mai sus: ALSCAL şi INDSCAL (în pachetul de bază), respectiv PROXSCAL (în modulul „SPSS Categories") (Meulman şi Heiser, 2001). Primele două proceduri sunt prezente şi în versiuni mai vechi ale SPSS-ului. Datorită răspândirii largi a programului în țara noastră, toate exemplele prezentate se vor baza pe aceşti algoritmi ai SPSS-ului. Totuşi, prin apelul la sintaxă, SPSS-ul pune la dispoziție şi alte tipuri de algoritmi.
- SAS-ul conține, printre altele, şi procedura MULTISCALE. Chiar dacă programul SAS nu este prea întâlnit în România, merită reținut că această procedură este revoluționară, fiind bazată pe estimarea rapoartelor de verosimilitate (în engleză „maximum likelihood estimation"). Acest fapt permite realizarea unor MDS confirmatorii şi nu doar exploratorii.

Iată o scurtă descriere a celor mai importante proceduri:

KYST (Kruskal, Young, Shepard, Torgerson) este unul din primele programe menite să realizeze scalare multidimensională în situații MDS metrice şi non-metrice de tip sintetic şi bazate pe relații de similaritate. Această procedură este cuprinsă în programul MDS(X).

MINISSA (Michigan-Israel-Netherlands-Integrated Smallest Space Analysis) este o altă procedură ce poate fi utilizată în condiții similare cu cele ale KYST. În versiunea sa metrică este denumit MRSCAL. Ambele versiuni sunt cuprinse în pachetul MDS(X). Avantajul programului față de KYST constă în

posibilitatea de a realiza o simulare pentru a compara soluția obținută (ex. numărul de dimensiuni obținute; valoarea stresului) cu cea rezultată în cazul unei distribuții aleatoare a punctelor obținută prin studii Monte-Carlo.

MDPREF (Multidimensional Preference Scaling) este un program adresat modelelor MDS metrice, sintetice, bazate în principal pe relații de tip unfolding, iar secundar pe relații de similaritate perceptivă. Pe lângă realizarea unui grafic cu puncte, este prezentat un vector care arată direcția preferințelor subiecților chestionați. Vectorul are săgeata îndreptată în direcția mediei preferințelor subiecților. Varianta non-metrică pentru această procedură este PREFMAP, care foloseşte cercuri şi puncte ideale pentru a sublinia preferințele respondenților. Ambele programe sunt incluse în pachetul de programe MDS(X).

TRISOSCAL (Triadic Similarities Ordinal Scaling) este un program adresat cazurilor de MDS non-metrice, sintetice, bazate pe relații de similaritate. Particularitatea se referă la modul de realizare a relației de similaritate. Nu se recurge la metode tradiționale de colectare a datelor. Metoda folosită este dată de comparații în triade, fiind întâlnită mai ales în domeniul constructelor personale, inițiate de Kelly în 1955 şi dezvoltate apoi de Fransella şi Bannister (1977). Triadele pot fi de două feluri: parțiale (alegeți din trei obiecte, două care sunt mai apropiate) sau totale (alegeți din trei obiecte, două care sunt cele mai asemănătoare; apoi alegeți din aceleaşi trei obiecte, două care diferă cel mai mult). Programul este inclus tot în pachetul statistic MDS(X).

MULTISCALE (Multidimensional Scale) este inclus în SAS şi, aşa cum am amintit, este singurul procedeu care are un caracter confirmatoriu. Într-un fel, MULTISCALE reprezintă în cadrul procedurilor MDS ceea ce analiza factorială confirmatorie reprezintă în cazul analizei factoriale.

ALSCAL (Alternating Least Square Scaling) este probabil cea mai răspândită procedură MDS. Ea poate fi utilizată atât în cazuri metrice, cât şi non-metrice, atât pentru relații de similaritate, cât şi pentru cele de tip unfolding. Totuşi, procedura este specifică datelor sintetice, fiind mai puțin eficientă în compararea diferențelor individuale (situații analitice). Programul este cuprins în numeroase pachete statistice precum SPSS Base System, SAS sau ViSta.

INDSCAL (Individual Differences Scaling) este o procedură aparte, fiind destinată nivelului analitic diferențial (priveşte diferențele individuale sau grupale). Procedura este inclusă în programe statistice cum ar fi SPSS Base System, SAS, sau MDS(X) şi este compatibilă atât pentru date metrice, cât şi non-metrice. Algoritmul poate fi utilizat pentru a reprezenta relațiile de similaritate dintre obiecte.

PROXSCAL (Proximity Scaling) este o procedură recent introdusă în pachetul adițional SPSS Categories, ce promite foarte mult. În mare, metoda conține două avantaje: poate fi aplicată pentru orice tip de date (metrice sau non-metrice), aflate în relații de similaritate sau preferențiale, vizând atât aspecte analitice (diferențiale, de replicare), cât şi sintetice. Procedura include astfel avantajele ALSCAL, INDSCAL şi IDEOSCAL şi, în plus, algoritmul de calcul utilizat permite o mai bună aproximare grafică a punctelor pe baza distanțelor dintre

obiecte. Principala limită a algoritmului rămâne imposibilitatea analizării datelor aflate în relații unfolding la nivel analitic. Altfel spus, nu putem surprinde diferențele individuale sau de grup în ceea ce priveşte preferințele pentru anumite obiecte decât prin apelul la artificii tehnice. O asemenea cale este introducerea unui punct ideal şi compararea acestuia cu fiecare dintre obiectele incluse în analiză pentru a afla gradul de similaritate sau de diferență a acestora față de punctul ideal.

8.2. Design-ul cercetării în scalarea multidimensională

Un design tipic implică rezolvarea a trei puncte fundamentale:

1. selectarea obiectelor analizate şi a numărului acestora;
2. metodele de colectare a datelor;
3. managementul datelor în utilizarea programelor statistice (ex. SPSS-ul).

Înainte de analiza fiecărui punct în parte, trebuie să amintim că un design tipic include, alături de datele privind gradul de similaritate / disimilaritate dintre obiecte, evaluarea acestora pe baza unor atribute externe (cu scopul de a uşura identificarea dimensiunilor latente), unele date demografice (cu scopul de a surprinde caracteristicile subiecților chestionați) şi, eventual, informații despre trăsăturile fizice ale obiectelor. Pentru a înțelege practic acest subcapitol, vom apela la un exemplu construit pornind de la structura cercetării întreprinse de Rhoades şi Arnold (1999). Fişierul aferent exemplului este „scalare similar.sav", iar calea de urmat în SPSS este „Analyze" - „Scale" - „Multidimensional scaling (PROXSCAL)". Dacă nu dispuneți de modul SPSS Categories veți utiliza calea „Analyze" - „Scale" - „Multidimensional scaling", după ce în prealabil ați copiat datele fiecărui grup „experți" vs. „novici", în două fişiere separate, deoarece modulul de bază al SPSS-ului nu permite compararea grupelor, dacă datele au fost introduse direct sub forma unei matrice de similaritate.

Context: În cercetările pe tema managementului conflictului au fost descrise şi clasificate o serie de comportamente şi modalități de reacție în situații conflictuale. Acestea se referă la stilul personal al fiecăruia de a răspunde, pornind de la reacții de tip nevrotic, până la strategii de acaparare a puterii. Concepte precum evitarea, compromisul, renunțarea, confruntarea, medierea, cooperarea, utilizarea forței fizice etc., sunt de notorietate în această problemă. Aceste categorii au fost grupate după nenumărate criterii taxonomice precum: orientarea spre sarcină vs. orientarea spre relații umane; active vs. pasive; unilaterale vs. bilaterale; constructive vs. distructive; individuale vs. de grup etc. Pornind de la aceste considerente, vom selecta ca obiecte supuse MDS următoarele zece reacții şi măsuri în situații de conflict: evitarea; compromisul; confruntarea; renunțarea; cooperarea; subminarea; manipularea; amânarea; arbitrajul şi medierea.

Pe baza acestora vom încerca să răspundem la următoarele întrebări:

1) care dintre criteriile taxonomice amintite mai sus au o relevanță mai mare? În termeni tehnici, care dintre criterii contribuie mai mult la stabilirea distanței psihologice dintre cei zece stimuli incluşi în studiu?
2) care este, din perspectiva eşantionului testat, ierarhia preferințelor de a reacționa în situații de conflict?
3) dacă împărțim eşantionul în două grupe: cu şi fără experiență în soluționarea conflictelor, există diferențe în ce priveşte modul de percepere a celor zece reacții incluse în studiu?

8.2.1. Selectarea obiectelor şi a numărului acestora

De cele mai multe ori nu putem include întreaga populație de stimuli dintr-o anumită categorie. Numai în exemplul precedent pot fi trecute în revistă mai mult de o sută de reacții la conflict, pornind de la cele de tip nevrotic (ex. inhibiție, frustrare), până la măsuri cu caracter global (grevă, război, embargo etc.). În selectarea stimulilor care vor face obiectul scalării multidimensionale este important să reținem trei aspecte, pe care le vom prezenta pe rând.

a) *Stimulii aleşi să fie cât mai reprezentativi* pentru categoria aleasă. Astfel, în studiile de marketing nu sunt incluşi toți stimulii, oricât de minori ar fi ei, ci se obişnuieşte să se aleagă stimulii cu o cotă de piață crescută, cu o diversitate cât mai mare în ceea ce priveşte atributele pe care le dețin. În exemplul nostru, se observă că stimulii aleşi sunt reprezentativi doar dacă facem referire la nivelul conflictelor interpersonale (Bogathy, 2002). Dacă vom ține seama de conflictele intrapsihice, intragrupale, intergrupale sau interstatale, vom observa că lipsesc multe dintre reacțiile tipice, posibile în astfel de situații conflictuale.

b) *Stabilitatea unei soluții este dependentă de raportul dintre numărul de stimuli incluşi şi numărul de dimensiuni selectate* de soluția optimă. Astfel, dacă anticipăm un număr mai mare de dimensiuni (ex. criterii de evaluare a unor produse), ar fi bine să includem un număr cât mai mare de stimuli. Au fost elaborate o serie de recomandări în acest sens. Jones şi Koehly (1993) enumeră două asemenea reguli. Prima susține că trebuie să avem între 4 şi 6 stimuli pentru fiecare dimensiune inclusă în soluții. Cea de-a doua, mai generală, sugerează ca numărul de dimensiuni extrase să nu depăşească valoarea (n-1)/4, unde n reprezintă numărul de stimuli (obiecte). În cazul nostru, având zece stimuli, ambele reguli indică o soluție stabilă reprezentată în plan (două dimensiuni). O soluție în 3 dimensiuni pare a fi mai instabilă, însă nu imposibilă. Această regulă nu are caracter obligatoriu, ci doar de recomandare, prin urmare, ea nu trebuie aplicată automat. Alegerea numărului de dimensiuni este dependentă în principal de indicatorii statistici obținuți şi de gradul de interpretabilitate a dimensiunilor. Mai mult, în situații cu un număr mare de obiecte, de exemplu 30, conform acestor reguli ar putea fi selectate optim până la şapte dimensiuni. Aceasta ar complica interpretarea rezultatelor, fapt ce a determinat o tendință puternică de a selecta,

cel mai frecvent, două sau trei dimensiuni. Alegerea numărului de dimensiuni va fi aprofundată într-un subcapitol ce va urma.

c) Un alt aspect, de care trebuie să se țină seama în selectarea numărului de stimuli incluşi, constă în *particularitatea MDS de a considera numărul de stimuli ca echivalent al numărului de subiecți*. Prin urmare, în utilizarea scalării multidimensionale pentru a identifica dimensiunile latente, N (volumul eşantionului) este dat de numărul de stimuli şi nu de numărul de subiecți. De aceea, se preferă un N care să conțină de la 8-9 stimuli în sus, cifra opt fiind minimul necesar pentru o soluție bidimensională.

Valoarea MDS optimă cuprinde 20 – 30 de stimuli. Numărul de stimuli aleşi este dependent şi de metoda utilizată pentru colectarea datelor, unele fiind nepotrivite pentru un număr mare de stimuli.

8.2.2. Metode de colectare a datelor

De la apariția tehnicilor MDS au fost dezvoltate o multitudine de metode de colectare a datelor, dintre care vom prezenta pe larg trei:

- compararea fiecărei perechi de stimuli;
- sortarea;
- ierarhizarea.

8.2.2.1. Compararea fiecărei perechi de stimuli

Este metoda cel mai frecvent întâlnită în MDS. Ea presupune compararea a câte doi stimuli. Numărul de comparații creşte foarte mult atunci când numărul de stimuli se măreşte, de aceea modalitatea este mai puțin potrivită pentru a fi utilizată la un număr de stimuli N>20. Numărul de comparații ce trebuie efectuate poate fi aflat cu ajutorul formulei [n(n-1)/2]. Astfel, pentru cei 10 stimuli din exemplul reacțiilor la situații de conflict, vor fi necesare 45 de comparații între stimuli perechi (ex. 1 cu 2; 1 cu 3; 1cu 4; ...; 9 cu 10). Dacă numărul de stimuli ar fi fost 12, utilizarea aceleiaşi metode ar fi necesitat 66 de comparații.

Din punctul de vedere al scalelor de măsurare, cea mai întâlnită modalitate este cea ordinală, de tip Likert, cu 5, 7 sau 9 diviziuni, în care alegerea unei valori capăt (ex. "1") ar însemna o foarte mare similaritate între obiecte, iar alegerea celuilalt capăt (ex. "7") semnifică diferențe foarte mari între obiecte. Iată două dintre cele 45 de comparații efectuate în problema mai sus menționată:

Figura 8.1. *Exemple de comparații perechi*

Conform acestor date, medierea, ca modalitate de răspuns la conflict, este văzută ca fiind mai asemănătoare arbitrajului. În schimb, între mediere şi confruntare există o distanță psihologică mai mare. Pentru a obține erori cât mai mici, înainte de a se trece la compararea efectivă a stimulilor se obişnuieşte să fie prezentată întreaga listă de stimuli, pentru a da posibilitatea respondenților să-şi formeze o impresie generală asupra gradului de proximitate dintre aceştia. Un program software dezvoltat special pentru acest lucru există şi în limba română. Denumit generic MultiScal, el a fost dezvoltat de firma SOFTAge S.R.L. printr-o finanțare dintr-un grant CNCSIS. Cei interesați să intre în posesia acestui program pot cere mai multe informații la adresa de e-mail prezentată în secțiunea introductivă.

8.2.2.2. Sortarea stimulilor

Sortarea stimulilor constituie o alternativă de colectare a datelor ce permite subiecților să grupeze stimulii în câteva categorii, bazându-se pe gradul de disimilaritate între date. Sortarea poate fi de mai multe feluri: singulară (subiectul alege o singură dată) sau multiplă (grupează obiectele de mai multe ori, după criterii diferite); restrictivă (un stimul poate fi inclus într-o singură categorie) sau nerestrictivă (un stimul poate fi prezent simultan în mai multe categorii); fixă (stimulii sunt grupați într-un număr prestabilit de categorii) sau liberă (pot fi formate oricâte grupe de stimuli).

Avantajul acestei metode constă în uşurința de a colecta rezultatele în cazul unui număr foarte mare de stimuli, depăşind astfel dezavantajul metodei comparației fiecărei perechi. Dezavantajul primar este dat însă de validitatea mai redusă a soluțiilor obținute pentru situațiile cu mai mult de trei dimensiuni. Mai mult, utilizarea unor proceduri de diferențiere interindividuală din seria INDSCAL a fost mult timp neaplicabilă în astfel de situații, iar soluțiile oferite de programe precum ALSCAL nu sunt întotdeauna corecte (ex. soluții degenerative) (Carroll şi Green, 1997).

În cazul propus, datorită numărului redus de stimuli (doar 10), este preferabil să nu utilizăm această metodă de colectare, din cauza dezavantajelor pe care le prezintă. În schimb, dacă am fi avut 40 de stimuli, metoda ar fi fost cu siguranță adecvată, deoarece apelul la compararea fiecărei perechi de stimuli ar fi presupus nu mai puțin de 780 de evaluări succesive. Cu scop pedagogic, în figura 8.2. sunt prezentate grupările celor 10 stimuli pe categorii obținute de la doi subiecți.

În aceste cazuri se observă că tipul de sortare utilizat pentru colectarea datelor este de tip singular (o singură sortare), restrictiv (un stimul nu poate fi simultan în mai multe categorii), şi liber (pot fi construite oricâte categorii şi pot fi alese oricâte elemente într-o categorie).

Rezultatele obținute se trec într-o matrice, în care atât pe linii. Cât şi pe coloane sunt trecuți stimulii prezentați. De fiecare dată când doi stimuli sunt sortați în acelaşi grup se primeşte un punct. În matrice, se trece indicele de disimilaritate

dintre doi stimuli pornind de la formula N-k, unde N reprezintă numărul total de subiecți investigați, iar k este numărul de subiecți care au sortat în aceeaşi categorie stimulii vizați. Astfel, dacă din zece subiecți chestionați, şase sortează în aceeaşi categorie evitarea şi renunțarea, indicele trecut în matrice, la intersecția dintre cei doi stimuli este patru.

Subiectul 1	Evitare Renunțare Amânare	Arbitraj Mediere	Cooperare Confruntare Compromis	Subminare Manipulare
Subiectul 2	Cooperare Arbitraj Mediere Compromis Confruntare	Evitare Renunțare Amânare	Subminare Manipulare	

Figura 8.2. *Exemplu de colectare a datelor prin metoda sortării*

8.2.2.3. *Ierarhizarea stimulilor*

Ierarhizarea stimulilor presupune ca subiecții să stabilească o ordine a preferințelor, acordând pentru fiecare stimul un rang. Spre exemplu, nota 1 revine stimulului preferat, nota 2 stimulului aflat pe locul secund în topul preferințelor, continuând până la nota N, acordată ultimului ales dintre stimuli. Această metodă este specifică modelelor MDS non-metrice, bazate pe relații de proximitate de tip unfolding. Ea nu oferă, cel puțin în SPSS, posibilitatea unor analize diferențiale pentru a surprinde deosebirile dintre subiecți sau dintre grupe ale eşantionului. În cadrul modelelor sintetice, rezolvabile prin SPSS, obiectele reprezintă variabilele cercetării, iar fiecare linie conține rangurile acordate în urma ierarhizării făcute de un singur subiect.

Alături de aceste modalități de colectare, întâlnite frecvent, există altele precum obținerea proximităților prin triade, compararea incompletă a perechilor, utilizarea aprecierii binare de tip asemănător vs. diferit etc. Mai multe informații sunt disponibile în lucrările lui Coxon (1982), Tsogo, Masson şi Bardot (2000).

Toate aceste modalități de prezentare a datelor intră în categoria celor obținute direct. Alături de acestea, se pot estima proximități între diferite obiecte pe baza unor indicatori derivați. Cel mai cunoscut indice derivat de disimilaritate este obținut pe baza valorii coeficientului de corelație după formula $(1-r^2)$. Se pot obține astfel reprezentări grafice care să ateste prin poziționare şi distanțare spațială, relațiile dintre mai multe concepte (obiecte). Din această perspectivă, MDS se apropie cel mai mult de modul de funcționare al analizei factoriale.

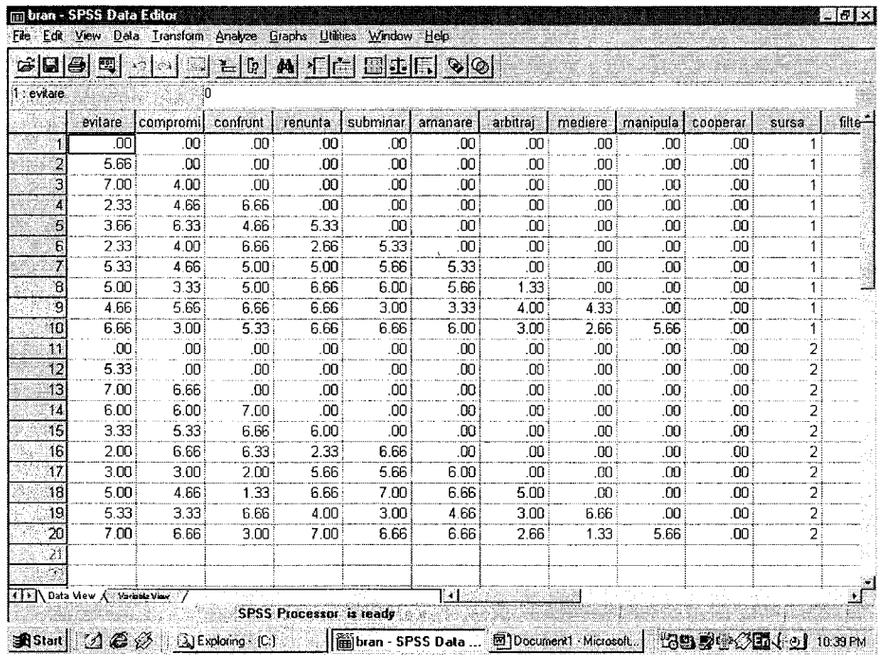
8.2.3. Managementul datelor colectate prin SPSS

Modul uzual de introducere a datelor în SPSS pentru a realiza o scalare multidimensională este sub forma unei matrice. Aceasta poate avea formă pătratică (număr egal de linii şi coloane) sau dreptunghiulară (rectangulară), poate

fi simetrică (relația dintre A şi B este aceeaşi cu cea dintre B şi A) sau asimetrică (AB diferit de BA).

Când avem mai mulți subiecți, se obişnuieşte să se facă media răspunsurilor la o anumită comparație. Astfel, revenind la problema cu reacțiile în situații de conflict, dacă avem opt subiecți testați, valoarea din matrice reprezentând gradul de disimilaritate dintre mediere şi confruntare este, de fapt, media răspunsurilor celor opt respondenți la comparația perechii respective. Dacă se consideră că este importantă analiza pe subgrupe de subiecți, vom construi mai multe matrice. Dacă dorim să realizăm o analiză diferențială la nivelul fiecărui individ, vom construi atâtea matrice câți subiecți avem.

Pornind de la exemplul dat, vom presupune că suntem interesați atât de o analiză de sinteză, la nivelul întregului eşantion, cât şi de diferențele existente între două grupe de subiecți: cu şi fără experiență în managementul conflictelor. Colectarea datelor prin metoda comparației fiecărei perechi de stimuli conduce la obținerea unor matrice simetrice şi de formă pătrată. Dacă datele ar fi fost selectate prin sortare, am fi obținut tot o matrice simetrică, de formă pătratică, fără însă a avea posibilitatea de a analiza diferențele subgrupale sau interindividuale. În schimb, dacă am fi obținut datele pe bază de ierarhizare a preferințelor, matricea ar fi fost asimetrică şi de formă rectangulară.


bran - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : evitare 0
SPSS Processor is ready
Start
Exploring - (C:)
bran - SPSS Data ...
Document1 - Microsoft...
10:39 PM

	evitare	compromi	confrunt	renunta	subminar	amanare	arbitraj	mediere	manipula	cooperar	sursa	filte
1	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1	
2	5.66	.00	.00	.00	.00	.00	.00	.00	.00	.00	1	
3	7.00	4.00	.00	.00	.00	.00	.00	.00	.00	.00	1	
4	2.33	4.66	6.66	.00	.00	.00	.00	.00	.00	.00	1	
5	3.66	6.33	4.66	5.33	.00	.00	.00	.00	.00	.00	1	
6	2.33	4.00	6.66	2.66	5.33	.00	.00	.00	.00	.00	1	
7	5.33	4.66	5.00	5.00	5.66	5.33	.00	.00	.00	.00	1	
8	5.00	3.33	5.00	6.66	6.00	5.66	1.33	.00	.00	.00	1	
9	4.66	5.66	6.66	6.66	3.00	3.33	4.00	4.33	.00	.00	1	
10	6.66	3.00	5.33	6.66	6.66	6.00	3.00	2.66	5.66	.00	1	
11	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	2	
12	5.33	.00	.00	.00	.00	.00	.00	.00	.00	.00	2	
13	7.00	6.66	.00	.00	.00	.00	.00	.00	.00	.00	2	
14	6.00	6.00	7.00	.00	.00	.00	.00	.00	.00	.00	2	
15	3.33	5.33	6.66	6.00	.00	.00	.00	.00	.00	.00	2	
16	2.00	6.66	6.33	2.33	6.66	.00	.00	.00	.00	.00	2	
17	3.00	3.00	2.00	5.66	5.66	6.00	.00	.00	.00	.00	2	
18	5.00	4.66	1.33	6.66	7.00	6.66	5.00	.00	.00	.00	2	
19	5.33	3.33	6.66	4.00	3.00	4.66	3.00	6.66	.00	.00	2	
20	7.00	6.66	3.00	7.00	6.66	6.66	2.66	1.33	5.66	.00	2	
21												
22												

Figura 8.3. *Print-screen SPSS cu matricea datelor*
SPSS este o marcă înregistrată.

Pentru simulare, vom presupune că avem un lot de şase persoane, împărțite egal în două grupe: experți şi non-experți în managementul conflictelor. În consecință, vom avea două matrice, fiecare având zece linii şi zece coloane. Dacă am fi interesați de modul în care întregul eşantion percepe cei zece stimuli, am avea o singură matrice cu zece linii şi zece coloane. Figura 8.3. prezintă imaginea fişierului „scalare similar.sav", după introducerea datelor în cele două matrice.

Se observă din imagine două matrice de tip 10X10, una aparținând subgrupului de non-experți (sursa = 1), iar cealaltă, grupului de experți în managementul conflictului (sursa = 2). Spre exemplu, valoarea de 4,66, existentă la intersecția coloanei 2 cu linia 4, reprezintă media evaluărilor subiecților din grupul non-experți vizând asemănarea dintre compromis şi renunțare. Se observă o corespondență între coloane şi linii (ex. atât coloana, cât şi linia a cincea se referă la subminare). În introducerea datelor este important să păstrăm ordinea liniilor pornind de la cea a coloanelor.

De asemenea, fiind vorba de date simetrice, există posibilitatea completării doar a unei părți din matrice. Această opțiune, cu completare sub diagonala principală, este şi opțiunea standard în PROXSCAL. Numai în cazuri de matrice asimetrice este necesară completarea întregii matrice. Din nefericire, aceste facilități sunt prezente doar în PROXSCAL, procedură inclusă în SPSS Categories. Pachetul de bază, care apelează la ALSCAL, permite efectuarea unei analize a diferențelor individuale numai în situațiile în care datele nu sunt direct proximități. În astfel de situații, datele sunt introduse normal (fiecare coloană reprezintă o variabilă, iar fiecare linie semnifică un subiect). Pornind de aici, se creează o matrice de date automată prin selectarea şi completarea opțiunii "Create distances from data". În cazul în care variabilele nu sunt apreciate cu aceeaşi unitate de măsură, se recomandă standardizarea lor inițială. Numai astfel de date pot fi supuse analizei diferențelor individuale prin INDSCAL, în pachetul de bază SPSS.

Dacă datele sunt construite prin metoda sortării stimulilor, vom introduce datele într-o manieră similară. Totuşi, este important să amintim că algoritmul de calcul nu este foarte bine pus la punct pentru a realiza analize individuale sau subgrupale prin această metodă. În consecință, nu se recomandă sortarea stimulilor pentru modele MDS analitice, ea fiind utilă pentru MDS sintetice, în situații cu un număr mare de stimuli, prezentați într-o singură matrice pătratică şi simetrică. O modalitate elegantă de a depăşi aceste neajunsuri este apelul la măsuri derivate. Astfel, dacă datele inițiale sunt de tip categorial, dihotomic, se poate apela la calcularea disimilarității după formula $1-\varphi^2$, unde phi reprezintă valoarea coeficientului de corelație pentru acest tip de date.

În sfârşit, dacă dorim să introducem date obținute în urma ierarhizării pe bază de preferințe a stimulilor, unica matrice introdusă va avea o formă asimetrică şi rectangulară. În cazul exemplului de față matricea va fi de tip 10X6, unde pe cele zece coloane se vor afla stimulii (variabilele), iar pe fiecare linie vor fi trecute rezultatele unui anumit subiect. Fiind şase subiecți chestionați, vom avea şase linii.

Rezultatele introduse în matrice se referă, de această dată, la rangurile acordate stimulilor respectivi în ordinea preferințelor [unde 1 = prima alegere, iar 10 (N) = ultima alegere]. Dacă unii subiecți nu pot diferenția între doi stimuli, se pot acorda ranguri egale (situație mai puțin dorită). Ierarhizarea (notarea) continuă apoi în condiții cunoscute şi întâlnite deja în cazul corelației Spearman a rangurilor. Cercetătorul are posibilitatea de a preciza, atât în acest caz, cât şi în situații de comparare a fiecărei perechi de stimuli, dacă acceptă sau nu egalitatea dintre ranguri. Iată print screen-ul SPSS 11.0 cu matricea datelor în urma ierarhizării stimulilor.

Matricea se citeşte foarte uşor: de exemplu, subiectul 4 clasează compromisul pe locul opt în ordinea preferințelor sale de a reacționa în situații de conflict. În schimb, subiectul 5 pune această reacție în topul preferințelor sale, pe locul al treilea.

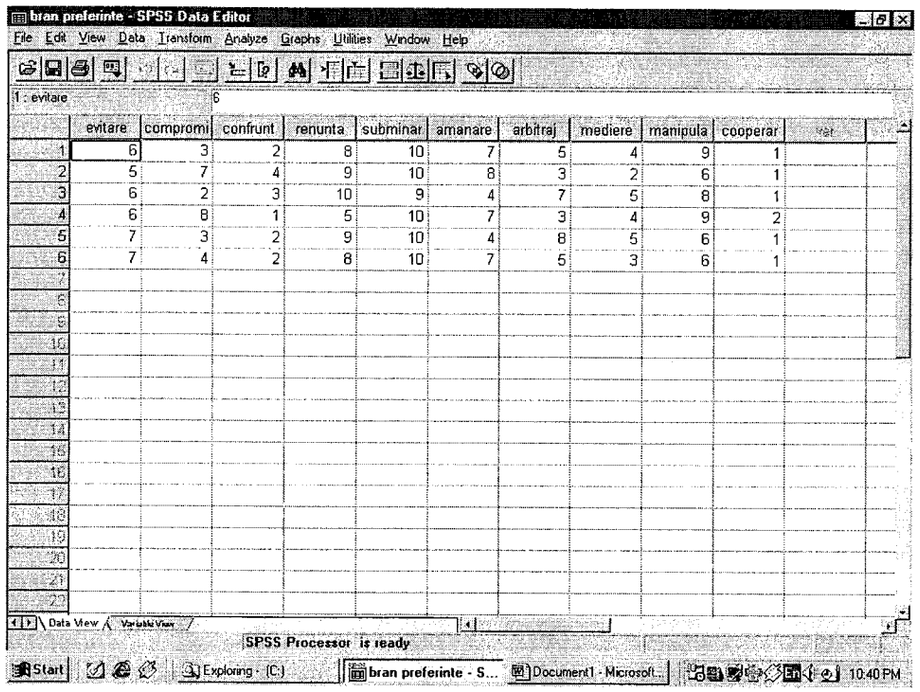

bran preferinte - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : evitare 6
evitare compromi confrunt renunta subminar amanare arbitraj mediere manipula cooperar
1 6 3 2 8 10 7 5 4 9 1
2 5 7 4 9 10 8 3 2 6 1
3 6 2 3 10 9 4 7 5 8 1
4 6 8 1 5 10 7 3 4 9 2
5 7 3 2 9 10 4 8 5 6 1
6 7 4 2 8 10 7 5 3 6 1
Data View Variable View
SPSS Processor is ready
Start Exploring - (C:) bran preferinte - S... Document1 - Microsoft... 10:40 PM

Figura 8.4. *Print-screen SPSS cu matricea rectangulară din fişierul „scalare preferinte.sav"*
SPSS este o marcă înregistrată.

Din păcate, SPSS-ul nu dispune de mijloace adecvate pentru analiza modelelor MDS analitice, unfolding. În schimb, datorită tipului de matrice dual (în engl. two-mode), există soluții alternative. Spre deosebire de matricele unimode

(one-mode) prezentate anterior, în care se face referire numai la stimuli (atât pe linii, cât şi pe coloane), matrice precum cea de mai sus conțin stimuli pe coloane şi subiecți pe linie. Situația permite modelelor extinse („unfolding") să reprezinte grafic atât stimulii, cât şi subiecții. Principiul de interpretare rămâne acelaşi. Cu cât un subiect este mai apropiat spațial de un anumit stimul, cu atât există o preferință mai clară pentru acesta.

8.3. Harta perceptivă. Reprezentarea spațială a distanțelor psihologice

În această secțiune vom discuta despre cum se stabileşte numărul de axe (dimensiuni) care contribuie cel mai mult la modul de aşezare a punctelor în spațiu. De asemenea, vom insista şi asupra modului de interpretare (identificare) conceptuală a acestora.

8.3.1. Selectarea numărului de dimensiuni din harta perceptivă

Scalarea multidimensională este în primul rând o tehnică descriptivă de reprezentare grafică a datelor. Cu excepția procedurii MULTISCALE nu există modalități de testare a ipotezelor legate de numărul de dimensiuni în care trebuie reprezentate spațial obiectele.

Totuşi, există mai multe indicii care pot fi utilizate în acest scop, printre care enumerăm: indicatorii statistici specifici (ex. stres sau DAF) (1), analiza grafică de tip scree plot (2), compararea indicatorilor statistici prin metoda Monte Carlo (3), gradul de stabilitate a soluției (4), uşurința de a reprezenta grafic soluția găsită (5), nu în ultimul rând, gradul de interpretabilitate a dimensiunilor (6).

Indicatori statistici în evaluarea numărului de dimensiuni

Gradul de potrivire sau nepotrivire a reprezentărilor spațiale pornind de la proximitățile exprimate de subiecți poate fi apreciat printr-o serie de indicatori statistici. Cel mai răspândit este conceptul de stres (raw stres sau raw phi). Au fost concepute mai multe formule de calcul, şi în consecință, mai multe forme de stres (Stres 1; Stres 2; Stres S). Din nefericire, valoarea acestora este dependentă de mai mulți factori, precum tipul de date (MDS metrice tind să ia valori de stres mai mari decât cele non-metrice) sau valoarea nominală a acestora (scalele de la 1 la 10 conduc la valori stres mai mari decât scalele de la 0 la 1) (Kruskal şi Wish, 1994). Din cauza acestor inconveniente, a fost obținută o formă normalizată de stres (N Stres), care face posibilă comparația dintre reprezentările grafice, indiferent de tipul de date. Aceasta are valori cuprinse între 0 şi 1 (Coxon, 1982). Pornind de la această normalizare, au fost propuse mai multe reguli de interpretare

a valorilor N Stres. În general, valorile mici ale stresului normalizat, precum cele cuprinse între 0,00 şi 0,15, indică prezența unei soluții adecvate, în timp ce valorile mai mari de 0,20 semnalează faptul că soluția găsită nu este potrivită (Kachigan, 1991; Kruskal şi Wish, 1994; Stalans, 1995).

Această interpretare nu poate fi făcută fără a ține seama de numărul de dimensiuni ale soluției, deoarece valoarea stresului tinde să scadă atunci când numărul de dimensiuni creşte. Mai mult, stresul este afectat şi de raportul dintre numărul de dimensiuni şi cel al obiectelor. Kruskal şi Wish (1994) arată că există situații cu o probabilitate de 50% de a obține un stres = 0 (reprezentarea a şapte obiecte într-un spațiu tetradimensional) sau un stres = 0,02 (aceleaşi condiții, dar într-un spațiu tridimensional). Ambele soluții suferă grav de lipsa stabilității. Chiar dacă am respecta raportul minim optim dintre numărul de obiecte şi numărul de dimensiuni, se ridică alte câteva întrebări. Ce facem dacă avem de ales între o soluție bidimensională (N Stres = 0,12) şi o soluție tridimensională (N stres = 0,08)? Ambele soluții satisfac criteriul valorii stresului. Care soluție ar trebui adoptată? Răspunsul la această întrebare ține de randamentul fiecăreia dintre soluții. Cu alte cuvinte, întrebarea devine: scade suficient de mult stresul pentru a accepta introducerea unei dimensiuni suplimentare?

Analiza grafică de tip scree plot

Analiza grafică de tip scree plot oferă un suport pentru a răspunde la această întrebare. Tehnica, introdusă de Cattell, este preluată din analiza factorială şi constă în realizarea unui grafic cu rol de suport. Pe abscisă (OX) sunt trecute dimensiunile posibile ale unei soluții, iar pe ordonată, valorile stresului (normalizat sau nu) pentru fiecare situație în parte. Interpretarea graficului porneşte de la ideea selectării acelui număr de dimensiuni de la care nu mai sunt aduse îmbunătățiri semnificative în evoluția stresului.

Să presupunem că avem 35 de obiecte introduse în MDS. Să ne reamintim că pentru a asigura stabilitatea soluției, numărul maxim de dimensiuni de care vom ține seama este opt (n-1)/4, iar numărul minim, unu. Fie valorile stresului (graficul din stânga al figurii 8.5.).

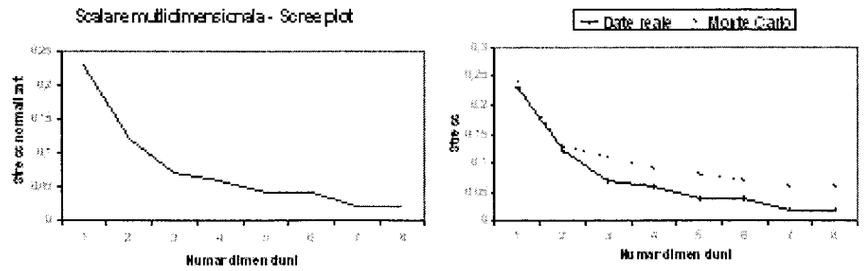


Figura 8.5. *Interpretarea grafică prin metoda „scree plot"*

Se observă că valoarea stresului normalizat scade cel mai mult (a se citi "soluția se îmbunătățeşte cel mai mult") când se trece de la o dimensiune (N stres = 0,23) la două dimensiuni (N stres = 0,12) şi de la două dimensiuni la trei dimensiuni (N stres = 0,08). După aceasta, scăderea stresului este prea mică pentru a merita să mai introducem o altă dimensiune. Grafic, această situație este reprezentată printr-o cotitură (punct terminus al unei descendențe accelerate şi început al unei descreşteri lente). Pentru a uşura interpretarea, Kruskal şi Wish (1994) propun o regulă suplimentară. Cu excepția unor cazuri bine argumentate teoretic, nu ar trebui să acceptăm acel punct de cotitură (implicit dimensiunea adiacentă), dacă valoarea stresului este mai mare de 0,10. Din acest punct de vedere, o soluție tridimensională pare a fi mai adecvată pentru graficul de mai sus.

Acest grafic poate fi realizat şi pe baza altor indicatori decât stresul. Printre aceştia, aş aminti DAF (dispersia explicată) sau coeficientul de congruență Tucker. Ambii indicatori, spre deosebire de stres, arată cât de bine se potriveşte reprezentarea grafică a punctelor cu afinitățile psihologice dintre concepte. Valorile lor, deşi cuprinse tot între 0 şi 1, au o semnificație inversă. Cu cât DAF sau indicele Tucker sunt mai apropiate de 1, soluția este mai potrivită. Din păcate, numai PROXSCAL-ul prezintă valorile acestor indicatori. Pachetul de bază SPSS prezintă doar valorile Stres 1, S Stres şi R, ultimul fiind un indicator echivalent pentru DAF.

Metoda Monte Carlo

O altă modalitate de selecție a numărului de dimensiuni este dată de metoda de simulare Monte Carlo. Aceasta poate genera randomizat soluții diverse de configurare spațială pentru numere diferite de dimensiuni. Prin compararea valorilor stres obținute prin simulare cu cele bazate pe date reale se poate alege numărul optim de dimensiuni. Soluția aleasă este cea care semnalează prima diferențe semnificative în ceea ce priveşte valorile stresului. În graficul din dreapta se observă că începând cu a treia dimensiune, soluția bazată pe date reale pare a fi semnificativ mai bună decât cea generată de configurații aleatoare.

Gradul de stabilitate a soluției

Un alt criteriu de sprijin utilizat în selectarea numărului de dimensiuni se referă la stabilitatea soluției. Efectuarea unei MDS pe 10 stimuli, precum în problema reacțiilor la conflict, duce doar la reprezentări grafice unidimensionale sau bidimensionale. Conform algoritmului descris anterior, ar fi fost nevoie de minimum 12 sau 13 stimuli pentru a putea lua în discuție o soluție stabilă tridimensională. În astfel de situații, cu doar două dimensiuni, analiza scree plot-ului nu-şi găseşte justificarea.

Uşurința reprezentării grafice a soluției

În alegerea numărului de dimensiuni este util să ținem cont şi de uşurința reprezentării şi a interpretării grafice. La bază, MDS este totuşi o tehnică explorativă de vizualizare a relațiilor dintre obiecte. Ce facem însă dacă soluția optimă are patru sau mai multe dimensiuni? Ar fi imposibil să reprezentăm grafic o asemenea situație şi ar fi complicat să interpretăm o asemenea configurație. De aceea, de foarte multe ori, cercetătorii au tendința de a selecta soluții bidimensionale sau tridimensionale, datorită avantajului oferit de posibilitatea reprezentării grafice. Totuşi, dacă există indicii clare că o soluție în patru dimensiuni ar fi mult mai adecvată decât una tridimensională, ar fi bine să o reținem, în ciuda dezavantajului legat de dificultatea reprezentării grafice.

Interpretabilitatea soluției

În fine, un ultim factor care influențează alegerea dimensiunilor ține de gradul de interpretabilitate a acestora. Acestui aspect îi va fi dedicat un întreg subcapitol. Până atunci, vom preciza că dintre două soluții relativ echilibrate pe baza criteriilor mai-sus amintite, este preferată aceea care oferă o mai bună interpretare a dimensiunilor şi implicit, a atributelor pe care subiecții le folosesc pentru a stabili proximitățile dintre obiecte. În ce priveşte numărul de dimensiuni selectate, mai rămân de discutat două probleme ce apar uneori în generarea soluțiilor MDS.

Prima, denumită **degenerarea soluției**, apare frecvent în condiții în care se fac MDS non-metrice utilizând obiecte care sunt grupate în câteva clustere naturale. Dacă întâlnim asemenea situații, interpretarea MDS se opreşte după identificarea acestor grupe de obiecte, fără a merge mai departe (ex. interpretarea dimensiunilor; analiza diferențelor interindividuale etc). Diagnosticarea unei asemenea situații este destul de uşoară. În astfel de cazuri, valorile N stres sunt foarte mici sau chiar zero (0,00 – 0,02). Obiectele sunt reprezentate grafic prin clustere depărtate spațial unul de altul. Dimensiunile obținute sunt greu sau deloc interpretabile, obiectele fiind configurate fără o logică anume. În plus, diagrama Shepard ia forma unei scări cu câteva trepte foarte mari. Totuşi, putem întâlni situații de soluții degenerate fără a observa în mod obligatoriu toate aceste simptome concomitent.

Cealaltă problemă ce poate apărea este **soluția locală minimă.** În acest caz, datorită unor deficiențe legate de procesul iterativ de generare a soluției, se pot obține valori destul de crescute ale stresului şi implicit, configurații mai puțin optime. Cu alte cuvinte, procesul iterativ se poate încheia înainte de a obține cea mai bună soluție. În sprijinul înțelegerii acestui proces vine imaginea clasică a unei persoane legate la ochi, care este pusă să găsească locul cu altitudinea cea mai joasă (valoarea stres cea mai mică). Dacă terenul este lin, persoana ar putea coborî pas cu pas până ar găsi locul cel mai jos posibil. Într-o asemenea situație,

soluția respectivă ar fi una corectă, iar în termeni statistici ar fi vorba de o echivalență între soluția locală minimă şi **soluția globală minimă**. Dacă terenul nu ar fi unul lin, ci unul accidentat, cu multe suişuri şi coborâşuri, persoana legată la ochi ar putea fi înşelată şi ar putea să declare locul cel mai jos posibil, o simplă înfundătură. Dacă ar fi luat-o într-o altă direcție, ar fi evitat această capcană şi ar fi ajuns, într-adevăr, în locul situat la cea mai joasă altitudine. Cum rezolvăm această problemă? Simplu, vom detecta mai multe puncte de start în călătoria persoanei. Dacă indiferent de acest punct, persoana ajunge în acelaşi loc sau într-un altul, cu altitudine apropiată, înseamnă că ne putem baza pe soluția obținută. Dacă există însă diferențe mari în ceea ce priveşte altitudinea punctului de sosire (valori etero-gene ale stresului), atunci întâmpinăm situația unei soluții minime locale. În astfel de situații, vom alege configurația corespondentă valorii N stres cea mai mică.

Situația în discuție poate fi recunoscută prin valori N stres destul de mari şi, uneori, prin evoluțiile paradoxale ale acestuia. Astfel, dacă în condiții normale stresul scade sau se menține constant o dată cu creşterea numărului de dimensiuni, în situații de soluții locale minime pot apărea cazuri ciudate în care stresul creşte, dacă se adaugă o dimensiune suplimentară (ex. o soluție tridimensională poate fi suspectată dacă valoarea stresului este mai mare decât în cazul unei soluții bidimensionale).

În astfel de configurații, interpretarea dimensiunilor şi a relațiilor dintre puncte se face dificil. Pentru a reduce sau elimina această problemă, PROXSCAL-ul din SPSS oferă posibilitatea de a alege configurații de start diferite (puncte de plecare a persoanei diferite). Oferta întâlnită în SPSS conține ca variante de start: Simplex (inițial punctele se află la distanță egală între ele), Torgerson (metodă clasică, propice mai ales MDS metrice), Random Single sau Multiple (care oferă una sau mai multe configurații generate aleator), Custom (permite configurarea inițială a punctelor pe baza unor rezultate anterioare).

Exemplu: Dimensiunile latente în evaluarea strategiilor de negociere

Din dorința de oferi mai multă claritate selectării numărului optim de dimensiuni, vom încheia prin a rezolva exemplul referitor la reacțiile în situații de conflict. Soluția care pare să se impună este aceea a unei reprezentări grafice bidimensionale. Iată şi argumentele în favoarea soluției MDS cu două dimensiuni:

- valoarea indicatorului N Stres era de 0,026, în condițiile în care nu au fost observate alte simptome ale unei soluții degenerative;
- valoarea suficient de mică nu indică nici prezența unei soluții locale minime;
- soluția bidimensională are suficient de multă stabilitate, dată fiind prezența celor 10 obiecte evaluate; de altfel, cele zece obiecte indică existența a doar două soluții alternative: MDS unidimensională sau bidimensională; prin urmare, argumentele care sprijină soluția tridimensională îşi pierd sensul, deoarece, o asemenea soluție ar fi instabilă din cauza numărului redus de obiecte (puncte);

- soluția poate fi reprezentată grafic cu uşurință; de altfel, reprezentările bidimensionale pe două axe Nord-Sud şi Est-Vest sunt şi cele mai uzuale în cazul hărților perceptive;
- cele două dimensiuni selectate sunt uşor interpretabile (vezi secțiunea 8.3.2.).

8.3.2. Interpretarea configurației MDS

Este timpul să reamintim o frază esențială prezentată în introducere. Proximitatea psihologică dintre obiecte poate fi reprezentată spațial sub forma unei distanțe între puncte. Cu cât punctele sunt mai apropiate unul de celălalt, cu atât obiectele sunt văzute de eşantionul testat ca semănând mai mult. La fel, cu cât punctele sunt mai depărtate unul față de celălalt, cu atât obiectele corespunzătoare acestora sunt percepute mai diferit. Distanțele dintre puncte nu trebuie absolutizate. Cu mici excepții, toate distanțele sunt doar măsuri relative. Aceasta înseamnă că nu putem trage concluzii de genul: obiectul A este de două ori mai apropiat de obiectul B în comparație cu distanța de la A la C, dacă observăm (ochio)metric o asemenea situație. Ceea ce contează este doar distanța relativă. Astfel, dacă distanța dintre punctele A şi B e mai mică decât cea dintre A şi C sau cea dintre B şi C, putem afirma doar că cele două obiecte A şi B par a fi percepute ca fiind mai asemănătoare decât oricare altă combinație. Pe baza acestor observații subiective putem grupa diferite obiecte în categorii. De exemplu, putem observa din figura 8.6 că subiecții tind să perceapă mai asemănător amânarea şi renunțarea ca forme de reacție la situații conflictuale decât, să zicem, amânarea şi cooperarea. Mai mult, am putea chiar identifica un cluster de reacții reprezentat de tripleta: amânare, evitare şi renunțare.

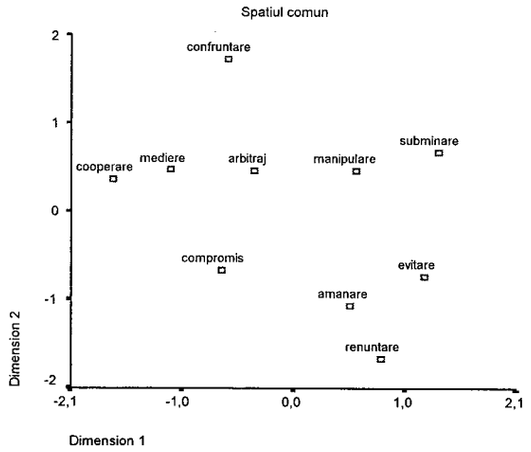


Figura 8.6. *Output SPSS cu reprezentarea spațială a distanțelor psihologice*
SPSS este o marcă înregistrată.

Interpretarea configurației MDS primeşte şi mai multă consistență, dacă identificăm dimensiunile pe baza cărora au fost plasate obiectele. Acestea nu sunt altceva decât criterii de judecată a gradului de proximitate între obiecte. Figura 8.6 prezintă harta perceptivă obținută în cazul problemei date.

Evaluarea subiectivă a dimensiunilor

O primă modalitate de interpretare (identificare) a dimensiunilor este de ordin subiectiv. Observați pentru câteva secunde distribuția punctelor de-a lungul primei dimensiuni. Găsiți vreun sens pentru modul în care sunt distribuite punctele? ... Foarte bine! Se pare că în partea stângă a primei dimensiuni se află situate reacții precum cooperare, mediere, compromis, confruntare, arbitraj. Ce au acestea în comun şi mai ales, ce le deosebeşte de punctele aflate în partea dreaptă, ca manipularea, evitarea, subminarea, renunțarea, amânarea? Cu excepția confruntării, se pare că medierea, cooperarea, compromisul sau arbitrajul, implică acțiuni bilaterale. În schimb, punctele din partea dreaptă par a fi mai degrabă acțiuni unilaterale. Din această perspectivă, am putea concluziona că prima dimensiune care contribuie la reprezentarea grafică a punctelor, şi implicit, la stabilirea asemănărilor sau a deosebirilor psihologice dintre obiecte, este axa reacții bilaterale vs. reacții unilaterale.

Într-un mod similar, am putea observa şi identifica configurația punctelor pe cea de-a doua dimensiune. Cum am putea denumi această axă, dacă la un capăt conține acțiuni energice, directe, de genul confruntării, iar la celălalt se regăsesc reacții precum compromisul, renunțarea, evitarea sau amânarea. Să fie vorba de un continuu al gradului de asertivitate? Să fie vorba de gradul de activism, pe o axă bipolară ce cuprinde reacții pasive vs. reacții active? Foarte probabil, mai cu seamă că spre mijlocul axei se observă reacții de nivel mediu în ceea ce priveşte gradul de asertivitate sau activism.

O astfel de interpretare subiectivă necesită cu siguranță multă intuiție şi sensibilitate. Trei pericole mari fac ca această metodă să nu fie prea iubită, în special de statisticieni. Primul pericol se referă la tendința oamenilor de a identifica paternuri de obiecte acolo unde de fapt nu există. Cel de-al doilea, se referă la posibilitatea efectuării unei rotiri ortogonale a dimensiunilor, schimbându-se astfel perspectivele de a privi punctele din grafic. Se cuvine aici să amintim că prin rotire nu se pierd informațiile esențiale legate de distanța dintre puncte. Să ne închipuim harta României. Indiferent dacă în urma rotirii vom localiza Timişoara sau Aradul în vest, sud, nord, sau est, distanța dintre ele se va păstra. În sfârşit, a treia şi cea mai consistentă critică, se referă la alte situații MDS decât cele bidimensionale. Astfel, este dificil să interpretăm configurația punctelor într-o reprezentare tridimensională (spațială) şi chiar imposibil, în situații mai complexe, cu patru sau mai multe dimensiuni.

Evaluarea obiectivă a dimensiunilor

Acesta este motivul pentru care au fost dezvoltate o serie de tehnici "obiective" de interpretare a dimensiunilor, indiferent de numărul acestora. Ele sunt denumite în jargon statistic adecvarea vectorială a atributelor (în engleză „property vector fitting"). În topul acestor măsuri se află regresia multiliniară. Principiul acestui ansamblu de tehnici este acela de a găsi atribute (trăsături) ale obiectelor proiectate care să coreleze puternic cu modul de configurare a punctelor şi cu coordonatele punctelor pe axele hărții perceptive. Pentru a utiliza tehnica adecvării vectoriale a atributelor este necesară colectarea unor date suplimentare. În funcție de modul în care este măsurată variabila dependentă se aplică regresia multiliniară, respectiv analiza discriminantului (ori regresia logistică). Dacă variabilele dependente constau în atributele complementare studiate, variabilele independente sunt reprezentate de coordonatele stimulilor (punctelor) pe axele hărții perceptive. Astfel, într-o situație MDS tridimensională vom avea trei predictori, iar într-o situație MDS bidimensională, ca în problema reacțiilor la situații de conflict, vom avea doar doi predictori.

Cum am proceda concret în acest exemplu?

Pasul 1. Identificarea variabilelor dependente care pot avea o semnificație psihologică, şi astfel, pot influența configurația spațială a punctelor. În exemplul dat am decis să evaluăm reacțiile la situații conflictuale prin prisma următoarelor atribute: măsuri pasive vs. active, unilaterale vs. bilaterale, individuale vs. colective, iraționale vs. raționale, distructive vs. constructive.

Pasul 2. Alegerea unuia dintre cele două moduri de măsurare a acestora (scală bipolară în şapte sau nouă trepte) în cazul regresiei multiliniare sau (opțiune dihotomică) în cazul analizei discriminantului sau a regresiei logistice. Vom alege pentru exemplul-model o scală bipolară cu şapte trepte. Astfel, pentru atributul pasiv-activ, se acordă un punct pentru o reacție considerată a fi pasivă, şapte puncte pentru una activă şi valori intermediare pentru nuanțe ale acestui aspect. Similar se procedează şi pentru celelalte atribute.

Pasul 3. Realizarea unei medii a răspunsurilor la un obiect pentru fiecare variabilă dependentă.

Pasul 4: Utilizarea coordonatelor fiecărui punct reprezentat spațial ca variabile predictor şi efectuarea analizei statistice corespunzătoare (vezi tabelul 8.1.). De remarcat că valorile coordonatelor pentru fiecare stimul pot fi salvate într-un fişier nou, prin selectarea opțiunii „common space coordinates", prezentă prin activarea butonului „output" din cadrul SPSS Categories, PROXSCAL.

După Kruskal şi Wish (1994), un atribut are importanță în interpretarea unei dimensiuni dacă, în urma analizei de regresie, îndeplineşte două condiții:

a) corelația multiplă dintre coordonatele punctelor şi atributul respectiv să fie foarte ridicată (R > .90, acceptabil însă şi peste .80); în situații limită (ex. R = .78), influența predictorilor trebuie să fie semnificativă statistic la un prag p < .01 sau chiar mai mic;
b) valoarea coeficienților standardizați beta (denumiți în cadrul MDS şi direcție cosinus) să fie cât mai apropiată de 1.

Tabelul 8.1. Valorile predictorilor şi ale criteriilor pentru fiecare atribut studiat prezente în baza de date „scalare coordonate.sav"

ATRIBUTE	DIM1	DIM2	Activ	Bilat.	Colectiv	Rational	Constructiv
evitare	1,233	-,772	1,50	2,66	1,50	4,17	2,75
compromis	-,683	-,703	3,75	5,60	3,25	5,85	3,50
confruntare	-,618	1,805	6,25	4,20	3,17	4,25	3,00
renunțare	,823	-1,746	1,33	1,83	3,00	3,80	2,50
subminare	1,370	,703	5,00	1,25	1,80	5,17	1,75
amânare	,539	-1,123	1,40	3,83	2,00	2,87	2,67
arbitraj	-,372	,485	4,00	5,00	5,17	4,50	3,33
mediere	-1,171	,490	5,00	5,50	5,00	4,33	4,17
manipulare	,591	,476	5,25	2,00	2,50	6,00	3,25
cooperare	-1,712	,386	5,50	6,17	3,50	4,83	5,83

Ce înseamnă îndeplinirea acestor condiții? Prima ne arată ce legătură există între configurația punctelor şi atributul respectiv. Cu cât corelația multiplă dintre predictori şi atribute este mai mare, cu atât atributul este mai semnificativ pentru modul de configurare a punctelor. Coeficientul de regresie standardizat (beta) ne arată cât de apropiat este atributul evaluat de sistemul de axe. Dacă beta pentru o anumită dimensiune are valoarea 1, atunci există o congruență totală între dimensiunea respectivă şi vectorul atributului respectiv, deoarece cosinus de unu este egal cu zero grade. Această situație ar constitui un indiciu extrem de puternic în favoarea denumirii dimensiunii respective prin apelul la acel atribut. Rețineți însă că este nevoie şi de o corelație multiplă ridicată. Celor mai puțin familiarizați cu aceste transformări trigonometrice le vin în ajutor manuale precum cel al lui Rogai (1983) care conțin tabele cu transformările necesare.

Rezultatele obținute în urma analizei de regresie pot fi sintetizate asemeni exemplului dat:

Tabelul 8.2. Analiza atributelor

ATRIBUTE	BETA DIM 1	BETA DIM 2	R
Pasiv - Activ	-0,226	0,830	0,939**
Unilateral - Bilateral	-0,989	-0,172	0,938**
Independent - Colectiv	-0,719	0,028	0,731
Irațional - Rațional	0,000	0,352	0,354
Distructiv - Constructiv	-0,926	-0,147	0,881*

** p < .001 * p < .01

Se observă că dintre cele cinci atribute evaluate, trei sunt puternic corelate cu modul de configurare a punctelor (R > .80). Mai mult, cele trei atribute sunt destul de apropiate de sistemul de axe. Astfel, vectorul unilateral - bilateral formează un unghi de doar 8º cu prima dimensiune (cos 0,989), iar vectorul distructiv – constructiv formează un unghi de 22° cu aceeaşi dimensiune (cos 0,926). Aceste unghiuri se măsoară din punctul de origine al hărţii perceptive, unde toate coordonatele iau valoarea zero. În sfârşit, vectorul pasiv – activ indică un unghi destul de ascuţit cu dimensiunea a doua a hărţii (cos 0,830 = 34º).

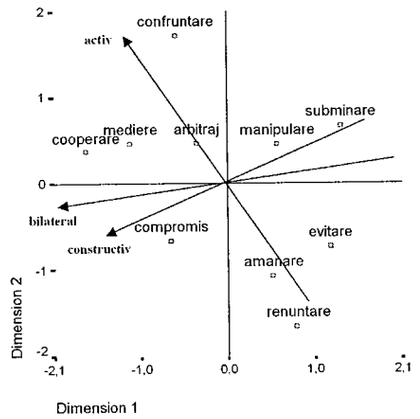


Figura 8.7. *Interpretarea dimensiunilor pe baza adecvării vectoriale a atributelor*

Direcţia vectorilor se poate stabili uşor, observând semnele coeficienţilor beta. Astfel, când ambii coeficienţi beta au semn negativ, săgeata vectorului va fi orientată în cadranul din stânga jos, corespunzător unor coordonate negative pentru ambele dimensiuni (cazul unilateral – bilateral, respectiv distructiv – constructiv). Similar, în cadranul de sus stânga, vor fi orientaţi vectorii care au beta negativ pentru dimensiunea 1 şi beta pozitiv pentru dimensiunea 2 (cazul pasiv-activ). Asemănător se procedează şi în alte cazuri. Din păcate, un astfel de grafic nu poate fi efectuat direct în SPSS. Alternative rămân utilizarea altor programe precum PROFIT [în cadrul pachetului MDS(x)] sau prelucrarea secundară a hărţii perceptive oferite de SPSS.

În figura 8.7. se observă tendinţa subiecţilor de a considera compromisul ca fiind o măsură constructivă şi bilaterală, confruntarea ca fiind, prin excelenţă, o măsură activă, medierea, cooperarea şi arbitrajul, ca reacţii bilaterale, constructive şi active. În acelaşi fel, am putea interpreta manipularea şi subminarea ca fiind distructive şi unilaterale, în timp ce evitarea, amânarea şi renunţarea sunt percepute ca reacţii pasive. Dincolo de această încărcătură a obiectelor în atributele respective, observăm că prima dimensiune ar putea fi denumită axa unilateral vs. bilateral, datorită unghiului redus dintre acest vector şi dimensiunea

în cauză. A doua dimensiune este mai greu interpretabilă, deşi am putea afirma cu un anumit grad de flexibilitate că este vorba de axa pasiv – activ (unghiul dintre vectorul respectiv şi a doua dimensiune fiind de 34º).

Aceasta nu este singura explicaţie posibilă. Dacă am roti ortogonal cele două dimensiuni cu aproximativ 20º în sensul acelor de ceasornic, am putea obţine o interpretare mai clară a dimensiunilor. Prima dimensiune ar putea fi dată de eficienţa reacţiei (măsuri constructive, cu consecinţe bilaterale vs. măsuri distructive sau beneficii unilaterale), iar a doua dimensiune ar constitui clar, axa pasiv – activ. Reţineţi, în fine, că relaţiile dintre vectori nu trebuie să fie ortogonale, în schimb dimensiunile, chiar după rotire, trebuie să formeze un unghi drept între ele. Mai mult, în cazul unor proceduri gen INDSCAL, dimensiunile nu trebuie rotite, deşi în practică mici rotiri pentru a maximiza interpretarea axelor sunt întâlnite (Jones şi Koehly, 1993). Spaţiul comun rezultat în urma configurării reacţiilor la situaţii conflictuale este un astfel de exemplu, în care rotirea axelor cu 20º oferă o interpretare mai clară a axelor.

8.3.3. Aplicaţii INDSCAL

Metodele MDS convenţionale nu pot epuiza întreaga gamă de întrebări şi probleme de cercetare. Uneori nu suntem interesaţi de obţinerea unei hărţi perceptive pe baza rezultatelor tuturor subiecţilor, ci dorim să observăm diferenţele dintre subiecţi sau subgrupe ale acestora. Reamintim că, în situaţia dată, subiecţii chestionaţi au fost împărţiţi în două categorii: cu şi fără experienţă în soluţionarea conflictelor. Oare nu există deosebiri între cele două categorii de subiecţi? Numai prin modele INDSCAL se poate răspunde la o asemenea întrebare. În urma aplicării unei astfel de proceduri vom obţine date pe care le vom prezenta pe rând.

(1) *Un spaţiu comun reprezentând configuraţia punctelor nediferenţiată pe categorii de subiecţi*

Graficul prezentat de noi în figura 8.6. este un astfel de exemplu, fiind o hartă perceptivă similară cu cea obţinută direct prin ALSCAL sau prin PROXSCAL identity matrix. Punctele vor fi la fel de distanţate, singura diferenţă care poate apărea fiind la nivelul axelor. Acest fapt nu constituie o problemă, deoarece dimensiunile pot fi rotite ortogonal în orice direcţie. Se constată totuşi, că modelele INDSCAL (singurele la care rotaţia nu este recomandată, fiind posibile doar mici ajustări) conduc la hărţi perceptive mai uşor interpretabile pe baza poziţiei iniţiale a axelor, în timp ce alte proceduri necesită, aproape permanent, rotiri ale axelor pentru a identifica corect dimensiunile obţinute.

(2) *Importanţa dimensiunilor pentru fiecare categorie de subiecţi* (în engleză „dimension weights" sau „saliences")

Pe baza acestora se observă eventualele diferenţe între subiecţi (grupe de subiecţi) în ceea ce priveşte gradul de adecvare a dimensiunilor selectate pentru fiecare categorie de respondenţi.. Pentru exemplul nostru, output-ul SPSS ar oferi următoarele rezultate:

Dimension Weights

Source	Dimension 1	Dimension 2
Fara experienta S1	,480	,459
Cu experienta S2	,512	,416

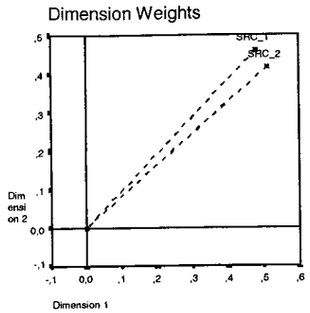


Figura 8.8. *Output SPSS cu analiza importanței dimensiunilor pe categorii de subiecți*
SPSS este o marcă înregistrată.

Nu există mari diferențe între modul de reprezentare a reacțiilor la conflict de către cele două categorii. Totuşi, se observă o uşoară tendință din partea persoanelor cu experiență de a acorda o mai mare importanță primei dimensiuni (distructiv vs. constructiv) în comparație cu cea de-a doua (WD1 = .512, față de WD2 = .416). Acest fapt pare normal, pentru lotul de persoane cu experiență în acest domeniu, ținând cont că diferențierea reacțiilor pe baza criteriului exprimat în această dimensiune este mai importantă în rezolvarea conflictelor. De notat că fiecare dimensiune selectată poate avea importanță diferită pentru fiecare categorie de subiecți, şi chiar pentru fiecare subiect în parte. Valorile pot fi cuprinse între 0 şi 1, unde zero reprezintă lipsa totală de importanță a acelei dimensiuni pentru categoria de subiecți respectivă, în timp ce 1 înseamnă exclusivitatea dimensiunii respective. De asemenea, între dimensiuni se pot stabili raporturi în ceea ce priveşte importanța acestora. În cazul de față, raportul importanței dintre cele două dimensiuni este 1,05 în favoarea primei dimensiuni pentru categoria celor fără experiență în soluționarea conflictelor şi de 1,23 tot în favoarea primei dimensiuni pentru categoria celor cu experiență. Aceste aspecte pot fi, de asemenea, reprezentate şi grafic (vezi figura 8.8.).

(3) *Spații individuale pentru fiecare categorie diferențiată*

Prin INDSCAL putem reprezenta grafic, cu scopul de a observa diferențele dintre acestea, atâtea hărți perceptive câte categorii au fost incluse în studiu. Pentru exemplul nostru, vor fi construite două asemenea hărți, deoarece am împărțit lotul în două categorii semnificative: cu şi fără experiență în soluționarea conflictelor. Interpretarea acestor grafice se face pe baza a două tipuri de informații: modificarea în configurația punctelor, respectiv gradul de creştere sau scădere a amplitudinii axelor. Iată cele două grafice obținute tot prin SPSS:

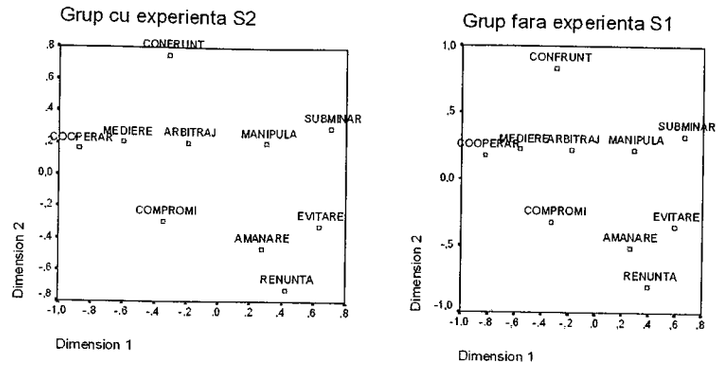


Figura 8.9. *Output SPSS 11.0 cu reprezentarea grafică a spațiilor individuale*
SPSS este o marcă înregistrată.

Configurația punctelor nu pare a fi modificată, dovadă fiind echivalența relației dintre punctele reprezentate în cele două grafice. Totuşi, am putea observa o uşoară tendință de creştere a dimensiunii 2 pentru grupul S1 şi a dimensiunii 1 pentru grupul S2. Cum s-ar interpreta acest lucru? Persoanele din grupa celor fără experiență tind să utilizeze mai frecvent dimensiunea pasiv-activ în gruparea datelor. Similar, persoanele cu experiență în soluționarea conflictelor tind să aibă o putere discriminativă mai mare în ceea ce priveşte dimensiunea distructiv – constructiv a obiectelor apreciate. Diferențele observate, bazate pe aspectele prezentate grafic în figura 8.9 sunt totuşi mici, interpretarea fiind forțată în cazul de față.

8.4. Modele de scalare multidimensională conexe

8.4.1. Modele MDS constrânse

În cercetare întâlnim frecvent situații în care avem date anterioare care pot fi integrate în configurațiile MDS. De exemplu, dacă majoritatea studiilor anterioare în domeniul reacțiilor la conflict ofereau ca dimensiune definitorie în construirea hărții perceptive gradul de distructivitate – constructivitate a reacției, atunci am putea constrânge harta perceptivă să conțină această dimensiune din start. Constrângerea duce automat la configurarea punctelor de pe hartă pe parcursul unei dimensiuni, astfel încât la un capăt să se afle reacțiile distructive, iar la celălalt reacțiile constructive. Pentru a constrânge stimulii de pe hartă este nevoie de configurarea inițială a acestora într-un alt fişier. Din păcate, prin utilizarea acestei metode toate dimensiunile selectate trebuie constrânse. SPSS-ul

oferă două posibilități în această direcție: identificarea directă a dimensiunilor, astfel încât fiecărei axe să-i corespundă un atribut şi realizarea unei combinații liniare între atribute în definirea unei dimensiuni. Prin acest tip de MDS nu mai trebuie interpretate dimensiunile însă rămân valabile celelalte aspecte ca: stabilirea gradului de adecvare a numărului de dimensiuni ales, interpretarea configurației punctelor, eventual observarea diferențelor interindividuale sau intergrup.

8.4.2. Modele MDS extinse

Aceste modele permit realizarea de hărți perceptive pornind de la preferințele subiecților chestionați. Dacă toate modelele discutate anterior arătau cât de apropiate sau de diferite sunt unele obiecte în percepția oamenilor, modelele extinse (unfolding) se finalizează prin hărți perceptive obținute pe baza preferințelor respondenților pentru obiectele studiate. De cele mai multe ori, astfel de date se prezintă sub forma unei matrice rectangulare, în care fiecare linie reprezintă rezultatele unui subiect. În plus, harta perceptivă va conține nu numai obiectele, ca în cazul datelor de similaritate, ci şi subiecții, poziționați grafic în funcție de preferințele exprimate.

În general există două abordări ale modelelor extinse (Hair, Anderson, Tatham, Black, 1998). Abordarea explicită constă în includerea unui obiect ideal între stimuli, urmând apoi procedura clasică de stabilire a gradului de similaritate dintre obiecte. Problema în cazul acestei abordări constă în tendința subiecților de a plasa acest obiect ideal, fie foarte departe de celelalte, fie foarte aproape de un alt stimul preferat. Mai mult, prin această metodă este greu să decelăm dacă subiecții au reuşit să aprecieze în comparațiile dintre obiecte, gradul de similaritate existent şi nu criteriul preferinței. Din aceste considerente este preferabilă abordarea implicită. Ea poate fi realizată pe baza unei analize interne sau externe.

Analize interne şi externe

Analiza internă implică ideea stabilirii directe a ierarhiei preferințelor fiecărui subiect pentru obiectele incluse în studiu. Datele sunt adunate într-o matrice rectangulară, care în cazul exemplului dat are 10 coloane (10 reacții la conflict studiate) şi 6 linii (6 subiecți chestionați) (vezi baza de date „scalare preferinte.sav"). În urma aplicării MDS corespunzătoare (versiune disponibilă şi în SPSS Base System), vom obține o hartă perceptivă ce include atât obiectele-stimul, cât şi subiecții. Mai mult, printr-o prelucrare secundară, similară celei prezentate anterior, se poate poziționa şi un vector al preferinței subiecților, rezultat dintr-o medie a răspunsurilor acestora. Interpretarea configurației punctelor pe hartă se face într-o manieră asemănătoare cu modelele MDS clasice. Astfel, dacă două puncte (fie ele stimuli şi/sau subiecți) sunt reprezentate grafic apropiat unul de celălalt, ele sunt percepute ca fiind apropiate şi din punctul de vedere al

preferințelor exprimate. Similar, distanțele mai mari dintre un subiect şi un anumit stimul indică o preferință scăzută pentru obiectul respectiv.

Analiza externă constă în stabilirea poziției unui punct ideal într-o configurație inițială construită pe baza similarității dintre stimuli. Pentru a realiza această analiză vom avea nevoie de colectarea datelor referitoare atât la similaritate, cât şi la preferință. Configurația inițială este construită pe baza similarității dintre stimuli. Pe baza acestora se pot stabili puncte ideale sau cercuri (zone) grafice încărcate cu grade diferite de preferință. Din păcate, SPSS-ul nu poate integra ambele proximități pentru a realiza analiza externă, deşi poate efectua separat ambele operații. Programul destinat acestui tip de analiză, denumit PREFMAP, se găseşte însă în pachetul MDS(X). Ținând cont de condițiile prezentate, vom exemplifica modul cum se interpretează o analiză internă, disponibilă şi prin SPSS.

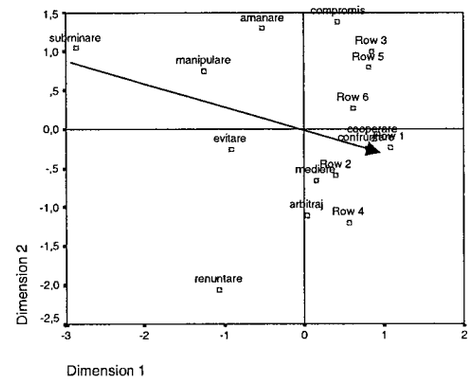


Figura 8.10. *Output SPSS 11.0 prelucrat cu reprezentarea grafică a observațiilor şi atributelor a MDS unfolding*
SPSS este o marcă înregistrată.

Din această reprezentare grafică se observă tendința subiecților 2 şi 4 de a prefera medierea şi arbitrajul ca modalități de reacție, precum şi a subiecților 1,3,5, şi 6 de a prefera cooperarea, confruntarea şi compromisul. Mai mult, există o tendință medie la nivelul întregului eşantion de a opta pentru aceste reacții la conflict, şi în special pentru cooperare şi confruntare. În schimb, se observă că subminarea, manipularea, evitarea, renunțarea sau amânarea nu stârnesc prea multe simpatii, prima fiind la nivelul întregului eşantion şi cea mai puțin agreată dintre toate formele de reacție.

Vectorul preferințelor a fost obținut într-o manieră similară cu cea descrisă la adecvarea vectorială a atributelor. Variabilele independente incluse sunt date tot de coordonatele spațiale ale obiectelor evaluate (nu şi ale subiecților), în timp ce variabila dependentă poate fi media (mediana) rangurilor acordate în ceea ce

priveşte preferinţele subiecţilor pentru fiecare obiect. În exemplul nostru, corelaţia multiplă obţinută între aceste elemente a permis constituirea unei axe preferinţă vs. nepreferinţă (R = .945; beta dim1 = -.958 şi beta dim2 = -.073) într-un spaţiu bidimensional (Stres 2 = .105).

8.5. Incursiune în opţiunile SPSS

SPSS-ul Base System (SPSS 11.0) oferă două opţiuni de introducere a datelor. În modelul clasic, fiecare coloană reprezintă o variabilă şi fiecare linie, rezultatele unui anumit subiect. În modelul matriceal, datele sunt măsurători medii ale proximităţii dintre obiecte. Din nefericire, a doua opţiune, nu este adecvată pentru modele de tip INDSCAL. Selectând butonul „shape" în "data are distances" avem trei opţiuni de matrice: simetrice (cazul comparării fiecărei perechi de obiecte), asimetrice (cazul testelor sociometrice) şi rectangulare (modele unfolding). Dacă optăm pentru "create data from distance", vom putea utiliza modelul obişnuit de introducere a datelor. Cele mai importante aspecte care trebuie selectate aici sunt modul de măsurare a datelor, cu variantele aferente fiecărei situaţii, respectiv standardizarea sau nestandardizarea variabilelor. Variabilele trebuie standardizate dacă nu sunt măsurate cu aceeaşi unitate de măsură. Apăsând butonul "Continue" revenim în fereastra principală, unde putem selecta tipul de model MDS. În cadrul MDS metrice putem alege variantele interval sau raport, în funcţie de datele avute, iar pentru modelele non-metrice, varianta ordinal.

De asemenea, putem decide între un model sintetic, ALSCAL (Euclidian Distance) şi unul analitic, INDSCAL (Individual differences euclidian distance). În plus, putem selecta numărul de dimensiuni dorite pentru care să obţinem indicatorii statistici de adecvare (ex. Stres). Deosebit de importantă este stabilirea tipului de condiţionalitate. Există trei opţiuni: „matrix" – pentru datele introduse direct ca distanţe în matrice; „row" – pentru modele MDS unfolding; „unconditional" – pentru cazurile "create distances from data". După alegerea modelului adecvat şi apăsarea butonului "continue", vom putea apăsa "options" pentru a selecta, în principal, modalitatea în care dorim să arate output-ul. Pe lângă aceste opţiuni, prin sintaxă putem apela la alte două modele MDS importante: GEMSCAL, pentru situaţii în care testăm aceiaşi subiecţi de mai multe ori şi ASCAL, pentru matrice asimetrice.

SPSS Categories oferă, începând cu versiunea 10.0, procedura PROXSCAL. Aceasta reprezintă un progres în MDS, fiind preferabilă utilizarea ei în comparaţie cu procedurile ALSCAL. Manualul SPSS Categories (Meulman şi Heiser, 2001) oferă mai multe amănunte legate de modul de utilizare a acestei proceduri şi, în plus, unele informaţii sunt similare cu cele prezentate anterior. Acestea sunt, foarte pe scurt, opţiunile ce trebuie selectate în principalele tipuri de scalare multidimensională.

Utilizarea modelelor analitice necesită formarea mai multor matrice, în timp ce MDS sintetice au ca design prezentarea datelor într-o singură matrice. În cazul exemplului nostru, am optat pentru un model echivalent INDSCAL (Weighted Euclidian). Paşii urmaţi pentru obţinerea rezultatelor prezentate în capitolul de faţă (cu excepţia MDS conexe) au fost:

1) introducerea datelor în SPSS sub forma unei matrice simetrice (vezi figura 8.3.);
2) selectarea opţiunii MDS PROXSCAL;
3) alegerea opţiunilor: „The data are proximities"; „Multiple matrix source"; „The proximities are stacked matrices across columns" şi apoi „Define";
4) introducerea variabilelor – reacţii la conflict în "proximities" şi selectarea variabilei referitoare la experienţa subiecţilor în situaţii de conflict în "source";
5) selectarea ca model a Weighted Euclidian (INDSCAL); „lower-triaungular matrix"; „ordinal untie tied observations"; „dissimilarities"; „within each source separately"; „minim2"; „maxim 3"; iar apoi „continue";
6) accesarea butonului „Plots" şi selectarea opţiunilor: „comon space"; „individual spaces" şi „individual weights" (opţional putând fi selectate şi celelalte grafice pentru a urmări gradul de adecvare a soluţiei);
7) selectarea altor opţiuni-output, în funcţie de aspectele de interes avute.

În fine, pentru MDS constrânse, cercetătorul poate interveni apelând butonul "restrictions", iar pentru modele unfolding, poate utiliza opţiunea "the proximities are in a single column".

BIBLIOGRAFIE

*** (2001). *SPSS Base 11.0. User's guide.* Chicago: SPSS Inc.

Bogathy, Z. (2002). *Conflicte în organizaţii.* Timişoara: Editura Eurostampa.

Carroll, J.D. & Green, P.E. (1997). Psychometric methods in marketing research: part II, multidimensional scaling. *Journal of Marketing Research*, 34, 193-204.

Coxon, A.P.M. (1982). *The user's guide to multidimensional scaling with special reference to the MDS(X) library of computer programs.* London: Heinemann Educational Books Ltd.

Fransella, F. & Bannister, D. (1977). *A manual for repertory grid technique.* New York: Academic Press.

Hair, J.F. Jr., Anderson, R.E., Tatham, R.L., şi Black, W.C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Jones, L.E. & Koehly, L.M. (1993). *Multidimensional Scaling.* În G. Karen & C. Lewis (Eds.). *A handbook for data analysis in the behavioral sciences: methodological issues* (p. 95-163). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kachigan, S.K. (1991). *Multivariate statistical analysis. A conceptual introduction.* (2nd ed.). New York: Radius Press.

Kruskal, J.B. & Wish, M. (1994). Multidimensional scaling. În M.S. Lewis-Beck (ed.). *International handbooks of quantitative applications in the social sciences. Vol.4 – Basic measurement* (p. 301-387). London: Sage Publications.

Meulman, J.J. & Heiser, W.J. (2001). *SPSS Categories 11.0*. Chicago: SPSS Inc.

Notelaers, G. (2002). *Multidimensional scaling. Experimenting with PROXSCAL*. Găsit la adresa http://homeusers.brutele.be/nova-inrct/Nederl/NavNOVA_r5_c1.htm#website.

Rhoades, J.A. & Arnold, J.A. (1999). The cognitive representation of responses to social conflict: the development of an integrative taxonomy. *The International Journal of Conflict Management*, 10(4), 360-384.

Rogai, E. (1983). *Tabele şi formule matematice*. Bucureşti: Editura Tehnică.

Rotariu, T., Bădescu, G., Culic, I., Mezei, E., Mureşan, C. (1999). *Metode statistice aplicate în ştiinţele sociale* (p. 254-271). Iaşi: Polirom,.

Stalans, L.J. (1995). Multidimensional scaling. În L.G. Grimm & P.R. Yarnold (eds). *Reading and understanding multivariate statistics* (p. 137-168). Waghington, DC: APA.

Tsogo, L., Masson, M.H., & Bardot, A. (2000). Multidimensional scaling methods for many-object sets: A review. *Multivariate Behavioral Research*, 35(3), 307-319.

van Deun, K. & Delbeke, L. (2000). *Multidimensional scaling*. Găsit la adresa: http://www.mathpsyc.uni-bonn.de/doc/delbeke/delbeke.htm.